DOCUMENT RESUME

ED 150 147 TM 005 961

TITLE	Educational Indicators: Monitoring the State of Education.
INSTITUTION	Educational Testing Service, Princeton, N.J.
PUB DATE	76
NOTE	106p.; Proceedings of the Educational Testing Service Invitational Conference (36th, New York, New York, November 1, 1975)
AVAILABLE FROM	Invitational Conference, Educational Testing Service, Princeton, New Jersey 08541 ($5.00)
EDRS PRICE	MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS	Academic Achievement; Awards; *Conference Reports; Cost Effectiveness; Data Collection; Developing Nations; Educational Objectives; Educational Practice; *Educational Quality; Educational Trends; Elementary Secondary Education; *Evaluation Criteria; Higher Education; *Information Dissemination; Information Systems; Measurement Techniques; Policy Formation; Prediction; Program Evaluation; Public Policy; *Quality of Life; Research Problems; Research Utilization; *Social Indicators; Sociocultural Patterns; Statistical Data; Trend Analysis
IDENTIFIERS	*Educational Indicators

ABSTRACT

The 1975 Educational Testing Service (ETS) Invitational Conference provided an overview of the social indicators movement, and the relationship between schooling and quality of life. The concept of educational indicators was discussed. Educational indicators are defined as statistics used to provide information regarding the status of particular variables, to compare trends, and to predict the future status of these variables. The ETS Award for Distinguished Service to Measurement was presented to Harold Gulliksen. The following papers were presented: The Social Indicators Movement, by Eleanor Bernert Sheldon; The Development and Use of Educational Indicators, by Dennis D. Gooler; Quality of Life as an Educational Outcome, by Stephen B. Withey; Measurement and Efficiency in Education, by Mancur Olson; Educational Indicators and Social Policy, by Wilbur J. Cohen; and Measures of Educational Outcomes in Developing Countries by Selma J. Mushkin and Bradley B. Billings. Priority Research Agenda on Educational Indicators was the subject of a panel discussion, involving all the speakers. (BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY
EDUCATIONAL TESTING SERVICE
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

ED150147

IM005 961

PROCEEDINGS OF THE 1975 ETS INVITATIONAL CONFERENCE

Educational indicators: monitoring the state of education

ETS

2

Educational Indicators: Monitoring the State of Education

Educational indicators: monitoring the state of education

PROCEEDINGS OF THE 1975
ETS INVITATIONAL CONFERENCE

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY

ATLANTA, GEORGIA
AUSTIN, TEXAS
BERKELEY, CALIFORNIA
EVANSTON, ILLINOIS
LOS ANGELES, CALIFORNIA
SAN JUAN, PUERTO RICO
WASHINGTON, D. C.

The thirty-sixth ETS Invitational Conference, sponsored by Educational Testing Service, was held at the New York Hilton, New York City, on November 1, 1975.

Chairman: David R. Krathwohl
Dean, School of Education
Syracuse University

Educational Testing Service is an Equal Opportunity Employer.

Copyright © 1975, 1976 by Educational Testing Service. All rights reserved.
Library of Congress Catalog Number: 47-11220
Printed in the United States of America

Contents

ix Foreword by William W. Turnbull

xi Preface by David R. Krathwohl

xv Presentation of 1975 Measurement Award to Harold Gulliksen

Morning Session

3 The Social Indicators Movement
Eleanor Bernert Sheldon

11 The Development and Use of Educational Indicators
Dennis D. Gooler

29 Quality of Life as an Educational Outcome
Stephen B. Withey

41 Measurement and Efficiency in Education
Mancur Olson

Luncheon Address

57 Educational Indicators and Social Policy
Wilbur J. Cohen

Afternoon Session

65 Measures of Educational Outcomes in Developing Countries
Selma J. Mushkin and Bradley B. Billings

81 Discussion: Priority Research Agenda on Educational Indicators
Eleanor Bernert Sheldon, Dennis D. Gooler,
Stephen B. Withey, Mancur Olson,
Wilbur J. Cohen, Selma J. Mushkin

ETS Board of Trustees, 1975-76

JUANITA M. KREPS, *Chairman*
Vice President, Duke University

ELIE ABEL
Dean, Graduate School of Journalism, Columbia University

ROBERT E. BELL
Principal and Vice-Chancellor, McGill University

JOHN W. HENNESSEY JR.
Dean, The Amos Tuck School of Business Administration, Dartmouth College

KURT M. HERTZFELD
Treasurer, Amherst College

ROGER W. HEYNS
President, American Council on Education

SIDNEY P. MARLAND JR.
President, College Entrance Examination Board

PETER MASIKO JR.
President, Miami-Dade Community College

GERTRUDE G. MICHELSON
Senior Vice President, Macy's

CHARLES D. O'CONNELL JR.
Vice President and Dean of Students, The University of Chicago

SISTER JOEL READ
President, Alverno College

WILSON C. RILES
Superintendent of Public Instruction, State of California

PAUL F. SHARP
President, The University of Oklahoma

STEPHEN H. SPURR
Professor, Lyndon B. Johnson School of Public Affairs, The University of Texas at Austin

WILLIAM W. TURNBULL
President, Educational Testing Service

ALBERT N. WHITING
Chancellor, North Carolina Central University

ETS Officers

PRESIDENT

William W. Turnbull

EXECUTIVE VICE PRESIDENT

Robert J. Solomon

VICE PRESIDENTS

Scarvia B. Anderson
David J. Brodsky
Jack R. Childress
John S. Helmick
Winton H. Manning
Samuel J. Messick
Charles E. Scholl
John Summerskill
Bernard L. Tchorni
E. Belvin Williams

ACTING VICE PRESIDENTS

Richard L. Burns
Robert E. Smith

TREASURER

Donald S. Grant

SECRETARY

Jane D. Wirsig

ASSISTANT VICE PRESIDENTS

L. Lynnwood Aris
Antoinette M. Orsi

ASSISTANT TREASURERS

Frank R. Henderson
Russell W. Martin Jr.

Foreword

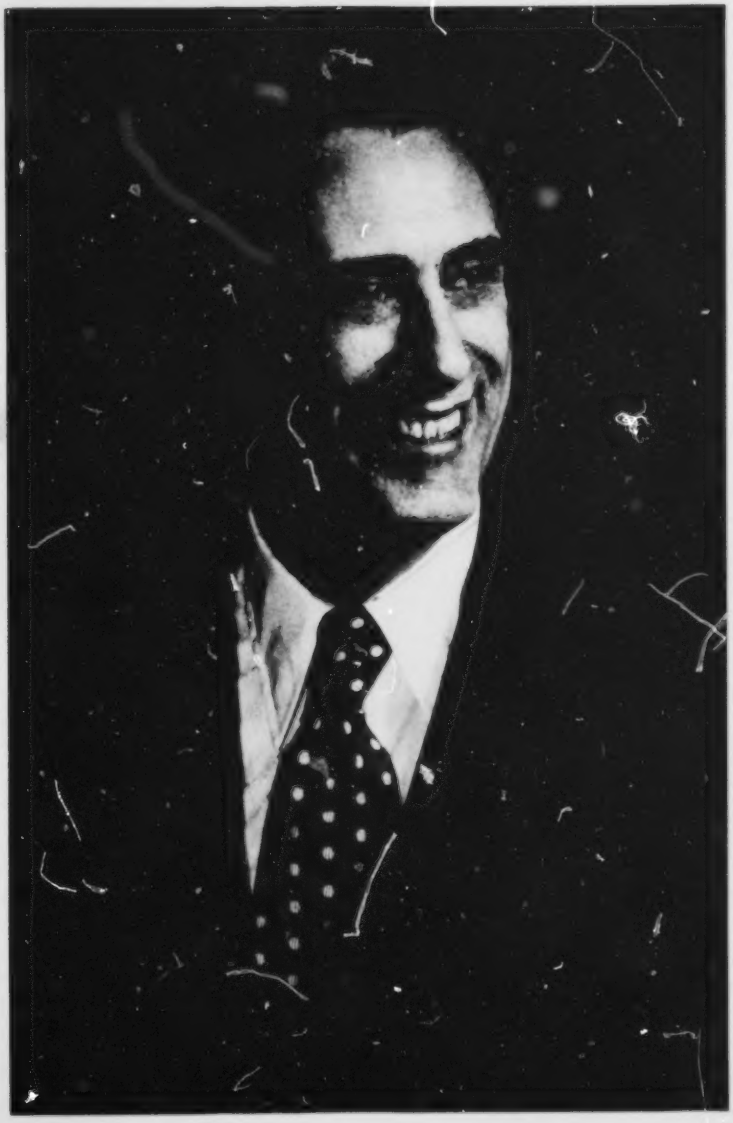
In recent years, there has been a growing movement to find some way of developing and using social indicators to take the nation's social temperature at regular intervals in the same manner as our gross national product and consumer price index measure its economic health. Such social indicators would be used to describe conditions and predict future trends in noneconomic areas of our society such as education or health.

The 1975 ETS Invitational Conference provided an overview of the social indicators movement, which draws upon a developing body of theory concerning the assessment of the functioning of large groups or systems. Speakers reviewed the origins and evolution of social indicators, their uses and limitations in our world today, and their potential for the world of tomorrow. What is the relationship between schooling and the quality of one's life and how can we use such data as an educational indicator? How can the information gleaned from social and educational indicators be used to guide sound educational policy? What international educational indicators have relevance for the United States?

The afternoon session featured a panel discussion in which all the speakers gave their views on the top priority research needs for further development of educational indicators. It was a very lively finale to a stimulating conference.

We are indebted to David Krathwohl who conceived of the theme of this conference and whose skill and diligence made it such a success. I should like to thank Dean Wilbur Cohen not only for a witty and engrossing luncheon speech but also for contributing valuable insights to the afternoon panel.

William W. Turnbull
PRESIDENT



David R. Krathwohl

Preface

Educational indicators is a topic whose time has come. Such measures are a natural outgrowth of the social accounting movement started by, among others, Dr. Eleanor Sheldon, author of our first paper. The efforts to construct indicators, such as those gathering momentum among the various social services, raise expectations for similar attempts in the education field, which repeatedly is diagnosed as needing change and improvement.

Massive federal legislation often has a great impact upon our schools, but key statistics are seldom available to help legislators, policy makers, and laymen understand in uncomplicated terms how education is changing. In the absence of other understandable indices, lower average Scholastic Aptitude Test (SAT) scores are interpreted by many as a drop in school quality. People ask whether the quality of education is commensurate with the financial investment in it; yet we lack indicators of both the investment and the resulting quality.

We must find better, simpler indices with which the decision makers and laymen who formulate federal, state, and local policy may more fully understand education. Perhaps such indices could be comparable to the gross national product, the cost of living index, or balance of payments, which help us understand our economic status. For example, as Dennis Gooler noted in his paper, Ralph Tyler has suggested that we calculate a "gross educational product."

The need for such measures has been reinforced by a mandate from Congress. The 1974 Educational Amendments to the Elementary and Secondary Education Act require the Assistant Secretary for Education to submit an annual report to Congress "... on the condition of education in the United States." The National Center for Education Statistics' recent publication, the highly useful *Condition of Education*, is a first response to that mandate. The need for indicators is also voiced in earlier government publications. For example, *Toward a Social Report*, compiled by Dr. Mancur Olson, whom we were fortunate to have as a participant at this conference, was one such governmental

effort. In addition, we have the increasingly successful efforts of the National Assessment of Educational Progress conducted by the Education Commission of the States.

In light of these activities and mandates, this conference on educational indicators seemed appropriate and timely. Eleanor Sheldon opened it by describing the broad context of social indicators. Dennis Gooler focused on the problems of defining indicators for education and proposed a concrete framework for a comprehensive system of such indicators. He also noted some of the likely problem areas.

Because achievement measurement in general, and National Assessment in particular, have been covered in past sessions, speakers at this conference addressed themselves to work being done on indicators in other areas of education. Mancur Olson examined some economic aspects of education, Steven Withey explored the relation of education to quality of life, and Selma Mushkin discussed indicators useful in international studies, giving particular attention to indices of self concept. To conclude the conference, the speakers formed a panel and examined some questions and research issues that bear on the future development of educational indicators.

Our conference speakers thus included individuals who helped initiate the social indicators movement, authors of comprehensive reports on education, and persons who have thought long and deeply about complex aspects of this issue. Therefore, we have reason to hope that this conference will contribute significantly to progress in the formulation of educational indicators and that work in this area can go forward with greater speed and clearer direction.

David R. Krathwohl
CHAIRMAN

EDUCATIONAL TESTING SERVICE

Measurement Award

1975

HAROLD GULLIKSEN

Harold Gulliksen

The ETS Award for Distinguished Service to Measurement was established in 1970, to be presented annually to an individual whose work and career have had a major impact on developments in educational and psychological measurement. The 1975 Award was presented at the conference by ETS President William W. Turnbull to Professor Harold Gulliksen with this citation:

> For almost half a century, Harold Gulliksen has made pioneer contributions to the solution of psychological problems by the application of mathematical methods. His work has been of marked significance for the advancement of knowledge in test theory, psychological scaling, and learning.
>
> For more than two decades, Dr. Gulliksen's *Theory of Mental Tests* was the leading text on test theory. He alone brought together, for the first time, information scattered throughout the literature, presented it with mathematical rigor and clarity, and added approaches and insights that gave mental testing a strong new impetus.
>
> His work in scaling provided new techniques for measuring more accurately those subjective states of individuals that are so important but yet so difficult to deal with in educational research.
>
> In learning, his research has embraced not only early studies of animals but also original mathematical theories of learning and transfer in human subjects.
>
> Professor Gulliksen was one of the founders of the Psychometric Society. His deep concern with the development of psychology as a quantitative rational science has also been reflected in his continuing role in guiding the editorial policies of the Society's unique and scholarly journal, *Psychometrika*, which he helped to initiate almost forty years ago.
>
> Perhaps his greatest contribution has been through his students. In addition to teaching at Ohio State University, the University of Chicago, the University of Iowa, and Princeton University, Professor Gulliksen served for almost twenty-five years as Director of the Psychometric Fellowship Program sponsored by Educational Testing Service and Princeton University. Throughout those years, his scholarship and integrity were an inspiration to successive generations of graduate students, many of whom are now distinguished mathematical psychologists at universities and research organizations throughout the world.
>
> For his many contributions in mathematical psychology and for his productive career as a scholar and teacher, ETS is pleased to present the 1975 Award for Distinguished Service to Measurement to Harold Gulliksen.

Previous Recipients of the ETS Measurement Award

1970 L.F. Lindquist
1971 Lee J. Cronbach
1972 Robert L. Thorndike
1973 Oscar K. Buros
1974 J.P. Guilford

Morning Session

The Social Indicators Movement

ELEANOR BERNERT SHELDON
President, Social Science Research Council

The past decade has witnessed an upsurge of interest in social measurement—largely in the form of proposals for the development of social indicators. The label is generally applied to measures of various social conditions and trends—both objective and subjective in nature—that is, measures of both external physical and social conditions and perceptions of these conditions. Several papers have described the origins and current outcomes of the wide range of activities that have come to be called "social indicators." Let us attempt a summary of that contemporary history.

Origins of the Movement

Social scientists, commentators, and policy makers brought the term "social indicators" into vogue during the mid-1960s. The term and allied phrases ("social reporting," "social accounting," "monitoring social change") emerged as a resulting confluence of many factors: an awareness of rapid social change and its consequent problems; the increasing diffusion of cost-benefit analysis and its possible application to the ever-increasing number and variety of domestic social programs; and a disenchantment among many social scientists who viewed the limitations of a structural-functionalism framework as impeding the re-emergence of interest in the analysis of social change.

In its early manifestations, the term "social indicators" and its synonyms became a rallying point around which many programmatic and evangelistic statements found legitimation and enthusiasm. The banners themselves (chiefly the banner of social indicators) became a symbol for a fashionable movement linking an ill-defined membership that shared few specific objectives,

leaned heavily on a few ambiguous symbols, and was led by articulate spokesmen. As in all movements, schisms developed, though this movement was so new it could not distinguish heresy from orthodoxy. All seemed to share a demand for social information, though questions pertaining to *what* information and for *what* purposes caused considerable splintering.

These early differences divided the movement into two main groups: 1) those seeking social information relevant to policy decisions; and 2) those seeking data to further our knowledge of the functioning of society and the measurement of social change.

The first group couched their indicator efforts in terms of the limitations of purely economic considerations in dealing with the problems of modern American society. Policies and programs aimed at the alleviation of problems sought data for planning, for implementation, and eventually for evaluation (8). These demands for social information were stated in terms of three uses: 1) the establishment of goals and priorities; 2) the evaluation of our social programs; and 3) the development of a system of social accounts that could provide guidance among many alternative interventions.

From the research community (the second major group) came demands for data that would be amenable to the measurement of change and the charting of social trends and that could enhance our capability for social prediction.

During the course of the early debate (8),

> it was pointed out to those seeking social information for policy and program purposes by the social science research community that their promises far exceeded any possibility for realistic attainment. It was noted 1) that priorities and goals are more dependent on national values than on assembled data; 2) that program evaluation necessitates the demonstration that the programs determine the outcomes (measured by indicators) rather than uncontrolled extraneous variables; and 3) that the essential theoretical prerequisite for the development of a system of social accounts—defining the variables and the interrelationships between them—was particularly deficient, if not completely lacking.
>
> The social science research community was reminded by the policy-analysis community that data are useful for planning and development of policies and programs; that

such data could be improved without awaiting the methodological and theoretical advances of the social sciences.

The Influence of Economists

Another debate, simultaneously waged and remaining extant (7, 10), emerges from the

> prominence of economists in the social indicators movement, and the fact that they saw the task as one of finding nonmarket measures of well-being, encouraged an emphasis on "non-economic" components of the quality of life and a vision of social indicators as measures of these components. The influence of economists was also responsible for the dominance of the imagery of the national economic accounts in discussions about social indicators. Both these ideas have served as rallying points for people in the social indicators movement. However, these ideas have provided an unproductive conceptual basis for the scientific work that is required for follow-through.
>
> A key problem is that the term "social" has been used in a residual sense to mean "outside the realm of economics." For example, Olson says: "The most notable limitation of the national income statistics is that they do not properly measure those 'external' costs and benefits that are not fully reflected in market prices." This is true enough, but the trouble begins when this approach to delineating the area of concern comes to define the realm of social indicators, as it does: "Ideally, what the national income statistics leave out, social indicators ought to measure, and a social report ought to assess."
>
> The implied homogeneity of the residual, of the "social" in this use of the term, does not exist. To suppose that it does is to guarantee confusion. And to suppose that the residual topics may, in principle, be represented by analogy to national income is to compound the confusion. Both points are well made by Leroy Stone (11):
>
> I have found that most people who use "social" in this sense [of noneconomic] fail to realize what a wide net they are casting, and how tremendously varied are the fish that the net will snag. . . We may think of social organization as being comprised partly by a network of interrelated and open sub-

systems. One of these subsystems is concerned with the production and distribution of wealth—we call it the economy. Examples of other subsystems are [the educational system and the system for acquisition and distribution of political power]. . .

The implied analogy to economic indicators development is actually useless; because nowhere in economic theory, as I understand it, are we faced with the problem of integrating information about such a wide variety of subsystems that are not demonstrably oriented toward any conceivable common goal that can be as fairly concretely identified as aggregate income.

The tendency to define the realm of the social as a residual, and what many regarded as an excessively aggregative, approach to indicator development was challenged by scholars who called for the resumption of detailed work to improve measures of change in various sectors of the society. The subject of measures of changes in social conditions had been discussed in one essay (Biderman's) in the Bauer book (1) and in a few of the papers in the *Annals* (4). An 800-page volume entitled *Indicators of Social Change: Concepts and Measurements* (9) was published in 1968 by the Russell Sage Foundation. This volume and other works contained detailed reviews of conceptual and measurement problems in the delineation of demographic, social structural, and other types of change in the United States. The data dealt with in these publications were of the hard variety, and in 1972, *The Human Meaning of Social Change* (2) was published, dealing with conceptual and measurement problems involving subjective data on public aspirations, expectations, and satisfaction.

Another approach aimed at the policy usefulness of social indicators may be referred to as "national goals accounting," for it attempts to give explicit definitions to national goals or social concerns. The government report, *Social Indicators 1973* (13), organizes its data by such social concerns and "widely held basic social objectives." A more fully explicated work is provided by Terleckyj in the recently published *Improvements in the Quality of Life* (12). Critics of this approach point out its limiting nature insofar as it is designed to select current social concerns and to provide output data relevant to them. The social system is divided into goal areas (thereby excluding all other areas), and outputs

from these systems are presented. The interrelations of inputs to outputs in the same goal areas and of outputs in the same and different areas are usually excluded. For instance, given that cognitive skill development is an output of the basic education system, achievement and attainment measures would be presented, though input measures (time spent in school) of the educational system and input and output measures of other related systems would be excluded. Among the criticisms, the critics note that output data cannot be adequately interpreted in the absence of system inputs.

Chief Purpose of Social Indicators

It is now generally agreed, though far from unanimously, that the chief purpose of social indicators is to comprehend what the main features of the society are, how they interrelate, and how these features and their relationships change over time. This is a realistic, albeit major, task and a long-term one. There are several aspects to it. It will involve work to improve the data base for social indicators (such as recently instituted surveys that develop new data on learning and the replication of questions asked in prior surveys to provide data for trend analysis). It will involve conceptual and methodological work such as recent developments in the subjective assessment of the quality of life and by work on models of social processes.

How do these tasks relate to education? Unfortunately, I am ill-equipped to provide definitive answers or even illustrative tasks. For discussion purposes, however, I offer the following comments:

1. *Improving the data base:* Relatively early (1969) in the indicators effort, the Russell Sage Foundation published a volume, *Indicators of Trends in American Education* (3). Its principal purpose was to bring together time-series data that indicate changing characteristics of education—viewed both organizationally and as a characteristic of the population. In assembling the data, Ferriss noted that

> the educational system offers a variety of statistical evidence in itself. One may be surprised to find so many time series

available in trends in education. Upon closer inspection, however, one may be equally amazed that so "literate" a system as Education, with a capital E, does not provide more evidence on many vital aspects of the system, such as the amount of learning that takes place or the qualifications of the teaching staff.

With respect to the latter, Ferriss provided time-series data on teachers with less than standard certificates and pointed out their weaknesses.

Since the time of these early illustrative observations (certainly not as a result of them), the National Assessment of Educational Progress has come into being and has issued a series of reports on educational attainments of young people. I am sure that most readers know more than I about the strengths, and particularly the weaknesses, of this endeavor, but surely it is an attempt to provide data on learning. It may warrant reconceptualization and refinement as well as methodological improvement.

Other improvements in the data base may also be forthcoming from the National Center for Education Statistics. Its most recent volume on *The Condition of Education* (14) deserves close scrutiny and critical comment for future improvement.

2. *Development of concepts and measures:* The concept of criterion referencing as distinct from norm referencing in testing seems to have blossomed in recent years, only partially as a result of the national assessment mentioned above. Again, I suspect that there is considerable debate about this but it, too, is a notion that warrants attention, definition, and methodological development.

3. *Social system models:* Many investigators are working on models of various aspects of change (6). Indeed, one early and persistent definition of social indicators placed them in the context of social system models. In 1971, Land (5) wrote: "I propose that the term social indicators refer to social statistics that . . . are components in a social system model . . . or of some particular segment or process thereof." Whether or not we adopt this confining definition, we need an approach that would interlock educational measures within the system with those from related fields (such as health), and that would relate these measures in cohort and social-historical time.

Toward the Future

Obviously, a vast amount of work remains to be accomplished. The movement is off the ground, less contentious, we hope, and certainly sufficiently secure to absorb alternative approaches that may prove to be mutually beneficial.

References

1. Bauer, R. (Ed.) *Social indicators.* Cambridge, Mass.: MIT Press, 1966.
2. Campbell, A., & Converse, P.E. (Eds.) *The human meaning of social change.* New York: Russell Sage Foundation, 1972.
3. Ferriss, A. *Indicators of trends in American education.* New York: Russell Sage Foundation, 1969.
4. Gross, B. (Ed.) *Annals of the American academy of political and social science,* 1967, pp. 371, 373.
5. Land, K.C. On the definition of social indicators. *American Sociologist,* Nov. 1971, 6, No. 4, 322-325.
6. Land, K.C., & Spilerman, S. (Eds.) *Social indicators models.* New York: Russell Sage Foundation, 1975.
7. Olson, M. Toward a social report, Part II: The plan and purpose of a social report. *The Public Interest,* 1969, 15, No. 86.
8. Sheldon E.B., & Land, K.C. Social reporting for the 1970s: A review and programmatic statement. *Policy Sciences,* 1972, 3, 137-151.
9. Sheldon, E.B., & Moore, W.E. (Eds.) *Indicators of social change: Concepts and measurements.* New York: Russell Sage Foundation, 1968.
10. Sheldon, E.B., & Parke, R. Social indicators. *Science,* 1975, 188, No. 4189, 693-699.
11. Stone, L.O. Comments on paper by E.L. Snider, presented at the Symposium on Social Indicator Research in an Urban Context, Ottawa, 1973.
12. Terleckyj, N.E. *Improvements in the quality of life: Estimates of possibilities in the United States, 1974-1983.* Washington, D.C.: National Planning Association, 1975.

13. U.S. Office of Management and Budget. *Social indicators, 1973.* Washington, D.C.: Government Printing Office, 1974.

14. U.S. Department of Health, Education, and Welfare, National Center for Education Statistics. *The condition of education.* Washington, D.C.: Government Printing Office, 1975.

The Development and Use of *Educational Indicators*

DENNIS D. GOOLER
Director of Research and Planning
University of Mid-America

Exactly five years ago yesterday, participants at the ETS Invitational Conference heard a series of papers grouped under the rubric "The Promise and Perils of Educational Information Systems." The Chairman of the 1970 Conference, Gene Glass, remarked in the conference *Proceedings* (8) that "the program of the 1970 Invitational Conference reflects a realization of the pervasive social consequences of the phenomenal inventions of twentieth-century psychometricians." Glass captured the general spirit of that Invitational Conference when he observed:

> The creation of an educational information system raises both hopes and fears. The promise of more informed decision making, which resides in these newly created systems, is quickly tempered in the minds of thoughtful men by the realization that these powerful inventions can be harmful if used carelessly.

Educational Indicators: Monitoring the State of Education, the title of this 1975 Invitational Conference, is clearly a central aspect in the development and use of education information systems. *Educational indicators* is the relatively new term being used to describe the means whereby such monitoring might occur. The term appears to be a derivation of the more global notion of social indicators, whose characteristics were outlined by Eleanor Sheldon a moment ago. A definition of educational indicators might be deduced from definitions of social indicators, as applied to education. More simply stated, educational indicators are statistics that enable interested publics to know the status of

education at a particular moment in time with respect to some selected variables, to make comparisons in that status over time, and to project future status. Indicators are time-series statistics that permit a study of trends and change in education.

There are innumerable elaborations upon this basic definition. For example, some argue that indicators must be statistics that can be disaggregated according to selected personal, geographic, or institutional attributes (14). Others argue that indicators must be of direct normative interest (4). In general, these elaborations call for statistics that are both meaningful and relevant to decision makers and consumers of education.

An examination of the subject of educational indicators might proceed according to three fundamental questions:

1. What must be known about education?
2. What will be done with what is known?
3. How can we gather needed information?

A discussion of educational indicators is most clearly associated with the question of how we can find out what we feel we must know about education. All of you are familiar with the problems of obtaining valid, reliable, and usable measures. Developing useful information systems, however, involves more than adequately solving technical measurement problems. Deciding *what* to measure and then *who* will measure it have always been difficult conceptual problems. These problems are fundamental to the development of an effective educational information system.

What is the condition of education? Consider the myriad ways in which education is described: number of persons graduating from high school; stabbings and drug use in elementary schools; senior citizens engaged in learning activities; violence related to busing; the strength of unions; why can't my children read?; empty school buildings; the New Math; inequalities in school finance; Ph.D.s driving cabs; American Nobel Prize winners. Bits and pieces of the story of education abound. How can we make sense of this incredible array of information? How can we know where we are in education, and where we've come from?

In many respects, education encompasses all thought, all avenues of human experience. To discuss education is to discuss much of what is known about mankind, both historically and in the future. And it is to discuss political processes, for education is

very much a part of the general tugging and pulling of ideas and values that must occur in human governance. The problem in establishing educational indicators is to simplify the complexity that is education to a form that is understandable while recognizing and honoring diversity and pluralism.

What Must be Known about Education?

As a first step, I shall attempt to set forth five categories of variables that address the question: What must be known about education? These categories might be seen as an organizing framework for the subsequent development of educational indicators, and thus, as an attempt to sort the various signals about the condition of education into a more usable form. Later, I will seek to describe the adequacy of our present information system.

The first category might be labeled *access*. Whom does education serve? In many respects, the question of access is the most global indicator of the condition of education. To understand education is, in part, to know who participates and who does not. Furthermore, it is important to describe the *kinds* of educational activities in which people engage. Many of these activities occur in formal institutions known as schools, but learning opportunities exist beyond these boundaries. Commercial and educational television and radio, for example, reach millions of people in a nonformal setting. Many adult education activities occur outside the formal educational system. These activities need to be recognized before a comprehensive picture of American education can be drawn. Access alone is not a sufficient, but may be a necessary, condition for equality of educational opportunity. Many people are repelled by the numbers game, but hard statistics on who does and who does not participate in education are indeed important statistics.

A second category, considered much less often, might be labeled *aspirations*. What do people want and need? To know the condition of education is to understand the extent to which education appears to contribute to fulfillment of personal, institutional, or societal needs and desires. Later on in this conference, the relationship of education to quality of life will be discussed. It is important to ask questions about that relationship. It is important

to know what people believe they want and what they believe they are capable of achieving. I wonder what it means for a man in his early fifties to declare that he is no longer capable of learning. What brings someone to this state? And what are the implications of this self-assessment for that individual's capacity to live his life as fully as possible?

Aspirations are extraordinarily difficult to define or to measure. And yet they seem so important, for they embody the idea of human will, which is so fundamental to achieving those things which are satisfying and desirable.

A third dimension along which the condition of education might be described concerns *achievement.* Most simply stated, what do people who participate in educational experiences learn from those experiences? To be sure, this is a different question from asking what people *know,* for people learn from so many sources. Nonetheless, it is important to understand what educators seek to teach, for such information indicates, at least in part, what people actually learn. Furthermore, it is important to understand what people can do with what they know. It may be important, for example, to ask not only "Can Johnny read?" but "Does he?"

For many years, education was talked about in terms of inputs. In the past decade, however, attention has shifted to focus on the outputs of education, most particularly the achievement of learners. Today, many people feel that measures of student achievement are the only appropriate measures of the condition of education. In fact, it would be difficult to imagine a statement of the condition of education that did not include an analysis of the achievement of participants in education.

A fourth category concerns *impact:* What happens to an individual who participates in educational activities? To what extent and in what ways is an individual's life changed because of participation in education? Most often, we look for relationships such as those that exist among schooling, occupation, and income. These are important dimensions, but there are many others. For example, it may be important to understand the effects of participation in education on feelings of self-worth or participation in community affairs or how a father treats his children.

It may also be important to understand the impact of education on general social, economic, and cultural systems. For example, what is the effect of beaming educational television programs via

satellite to societies that heretofore have had little technology available to them? What does it do to their culture? These are difficult questions, the answers to which may be found only in long-term studies. And yet the condition of education will not be fully understood until the impact on individuals, institutions, and even society can be ascertained and tracked over time.

Finally, the condition of education might be examined according to the *resources* available to education. What is spent on education? What kinds of material and physical resources are available to education? What is the quality of human resources available to those who participate in an educational activity? What is the ratio of cost to benefit? How do the type and quality of current investments in human capital influence the availability of future resources?

Figure 1
Categories of Educational Indicators

ACCESS
- How many and what kinds of people participate in educational activities
- Retention rates in educational activities
- Catalog of existing/available educational activities or services

ASPIRATIONS
- Description of needs and desires of various kinds of people
- Individual self-assessments of personal capabilities
- Description of institutional goals

ACHIEVEMENT
- What people know, do, and feel
- What people have earned (degrees, diplomas, certificates)
- What is taught

IMPACT
- Consequences of having schooling
- Impact of education on social/economic/cultural systems
- Consequences of not having schooling

RESOURCES
- Capital, personnel, and materials expenditures
- Quality of human resources
- Cost to benefit/effectiveness ratios
- Quality of educational climate
- Time

Education must constantly compete with other kinds of programs for limited resources. It may be important to examine over time the proportional growth or decline of resources available to the educational enterprise. This becomes particularly important as one seeks to affix a measure of accountability to education. If one is to be accountable for something, one presumably must have the resources to do that something.

If these five categories were acceptable as representing the basics of what must be known about education, and if indicators were available or could be developed for each category, we could then describe the characteristics and numbers of people participating in some way in education, what they want and need out of life and from education, what people know and how they are taught, the impact of education on individuals and institutions, and the resources available to education. This would be a comprehensive picture indeed—not complete but quite comprehensive.

The most apparent characteristic of much of this information, which is summarized in Figure 1, is that it is relatively simple. Too simple, some will argue. Education is a complex task, and indicators err on the side of portraying this complexity inadequately. The same argument is made by those working intimately with any project or idea when someone seeks to describe their work. Bob Stake (15, pp. 65, 66) spoke to this same idea in an address on National Assessment:

> National Assessment is an effort to simplify and bring within reach of our understanding the robustness of education in this nation. It cannot help but be an oversimplification. But we know that every index number, every graph, every word of prose is an oversimplification. We have no choice but to create simple things to stand for complex ones. Our curiosity, our desire to command our destiny, demand it. We are human beings. We will not be persuaded that it is wrong to define, to symbolize, to model, to measure.

Part of the difficulty may stem from the fact that experts see more complexity in things than do laymen. Thus, professional educators may find indicators more difficult to use, or less meaningful, than do consumers of educational services. This, however, should not preclude the development of indicators; the consumer is important.

How Adequate Is Our Present Information System?

How adequate in the eyes of producers, consumers, and regulators of education is our present information system *vis-à-vis* this framework? The question is complex, largely because there is so much extant data but so little systematic effort to put it together. In the analysis to follow, I will focus on a relatively few attempts to aggregate and analyze Census data at the national level. In so doing, I do not adequately reflect massive efforts such as Project Talent (7), the work of the International Association for the Evaluation of Educational Achievement (13), reports of commissions such as the Carnegie Commission on Higher Education (2), and countless others. All such efforts supply valuable insights into the condition of education but are not treated in depth here.

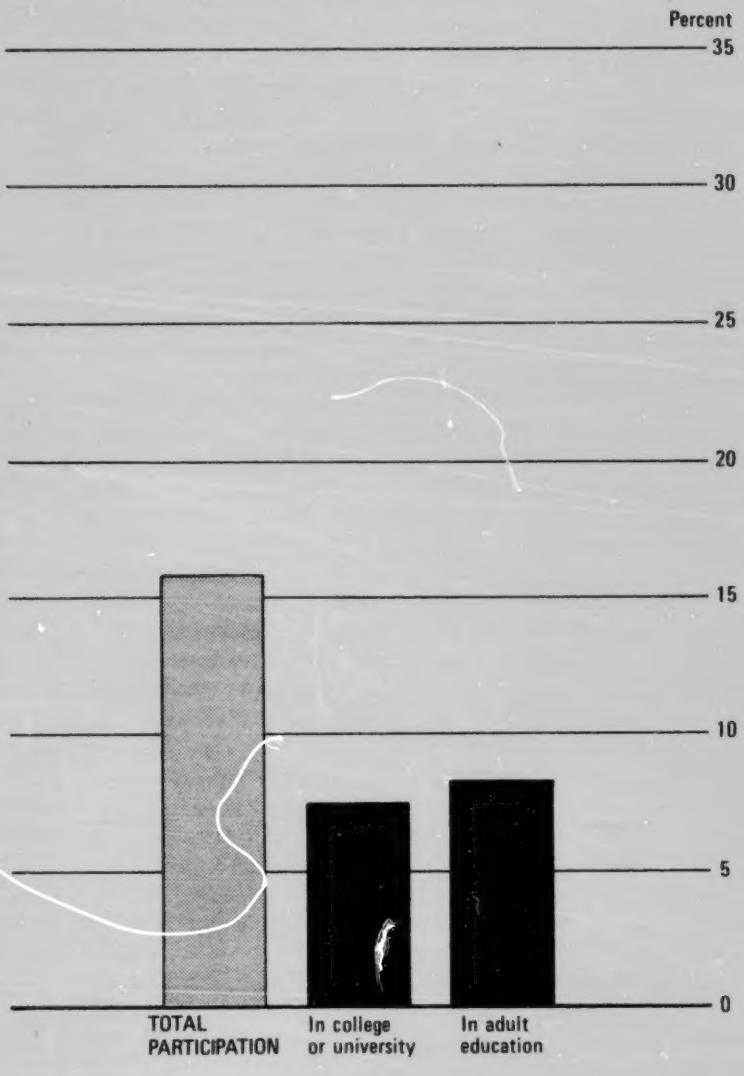
We appear to know a great deal about access, if participation in formal educational activities can be taken as one indicator of access. Data reported by the National Center for Educational Statistics (11), by the Office of Management and Budget (12), and analyzed in depth by, for example, Duncan (5), give a rather extensive picture of who participates in what kind of formal educational programs. These data, largely from the Census, can and will be collected regularly to show changes in participation in education over time.

An examination of these reports reveals participation data disaggregated by type of institution, sex, race, age, and family income among other variables. Retention rates are also reported, primarily in secondary education. We have some idea of how many adults participate in what kinds of educational activities. Figure 2, for example, is taken from *Social Indicators 1973* (12). Note the simple rendering of this statistic.

What is less apparent from existing information is how accessible educational opportunities are *perceived* to be or how many people participate in nonformal educational activities. The conditions of access, such as available transportation or facilities for the handicapped, may influence people to attend educational programs. Despite the absence of data about these specific issues, there is a substantial amount of information available on who participates in education and, thus, on its accessibility.

The National Center for Educational Statistics (NCES) has been and will continue to be involved in systematic data-collection

Figure 2
Educational Participation of Adult Population: 1969
(From *Social Indicators 1973*, p. 86)



18

efforts to address the access question. The Elementary and Secondary General Information System of NCES, for example, is the primary vehicle for gathering quantitative data on students. New efforts in postsecondary education are also under way at NCES.

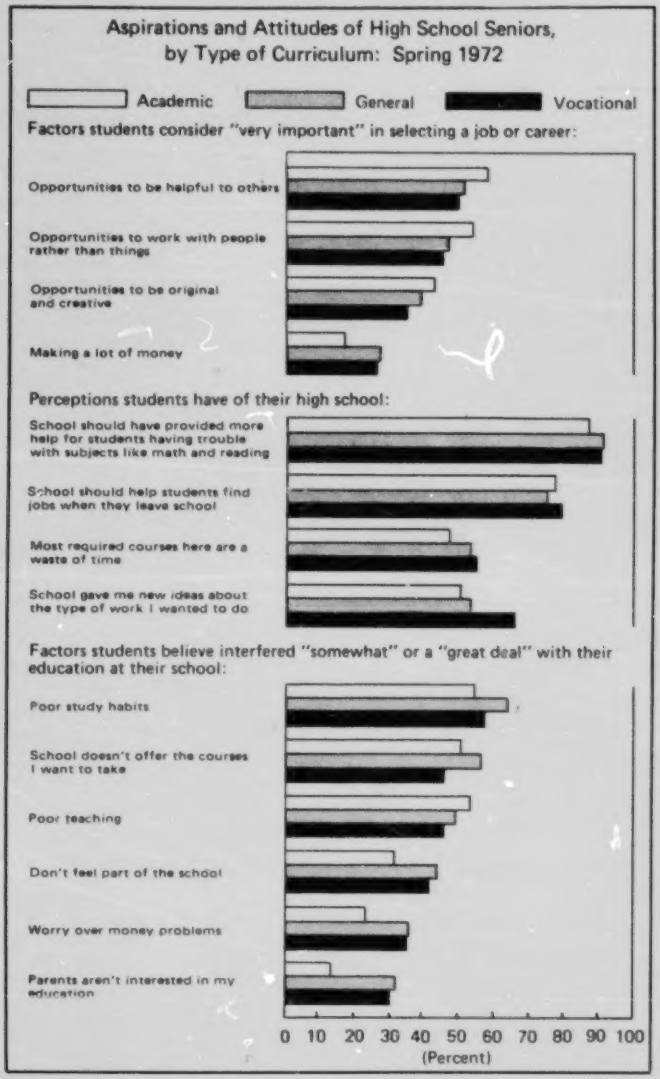
It may be well to pause here to point out that these statistics do not tell us whether *enough* access is provided or whether the educational system is just. Such judgments entail comparing what exists with some standard or expectation or comparing present with previous status. Describing the *condition* of education is different from judging the *well-being* of education.

When one considers our understanding of aspirations, the picture does not seem as bright. There are countless studies of attitudes toward education, degree aspirations, attitudes toward self, and the like, but apparently very little has been done to regularly and systematically obtain a general picture of the hopes, fears, and aspirations of people. For example, *Social Indicators 1973* (12) contains no data on the educational aspirations of citizens or, indeed, on aspirations of any kind. The *Condition of Education 1975* (11) includes one bar graph containing information roughly approximating the concept of aspirations. The bar graph is reproduced on page 20 as Figure 3 to give an example of the kind of information being reported.

Plans are under way, however, to develop a better data base on aspirations. NCES has plans to continue its Longitudinal Study of Educational Effects, which will provide data on the extent to which career plans and aspirations persist over time. Sidney Micek and his colleagues at the National Center for Higher Education Management Systems (NCHEMS) include in their higher education outcome measures system a number of items related to student aspirations. Gallup and Harris polls continue to contribute to an understanding of what people want and how well they assess their own capacities to achieve their goals. Work such as that included in the volume by Campbell and Converse (1) is useful in understanding aspirations and the human meaning of social change.

Data about aspirations is regarded by some as being too soft for inclusion in an information system. Others say participation in an activity can be defined as an indicator of aspirations. I contend that the quality of education must be assessed, at least in part, by the extent to which individual or collective aspirations are clarified

Figure 3
Aspirations and Attitudes
(From *The Condition of Education 1975*, p. 31)


Aspirations and Attitudes of High School Seniors, by Type of Curriculum: Spring 1972
Academic
General
Vocational
Factors students consider "very important" in selecting a job or career:
Opportunities to be helpful to others
Opportunities to work with people rather than things
Opportunities to be original and creative
Making a lot of money
Perceptions students have of their high school:
School should have provided more help for students having trouble with subjects like math and reading
School should help students find jobs when they leave school
Most required courses here are a waste of time
School gave me new ideas about the type of work I wanted to do
Factors students believe interfered "somewhat" or a "great deal" with their education at their school:
Poor study habits
School doesn't offer the courses I want to take
Poor teaching
Don't feel part of the school
Worry over money problems
Parents aren't interested in my education
0 10 20 30 40 50 60 70 80 90 100
(Percent)

and approached. The data may be soft if what we use are self-reports of aspirations, but they are nonetheless valuable and need to be studied seriously in a responsive educational system.

With respect to achievement, the most commonly used index is degrees or diplomas earned, or, at least, number of years of schooling. Now most educators recognize the amount of faith required to accept possession of a diploma or degree as a meaningful representation of how much or what people know, particularly in light of the plethora of kinds of diplomas available. The variation across schools offering diplomas or degrees is considerable. Nonetheless, the attainment of a diploma or degree *is* a kind of achievement and can be treated as such.

Some Indicators of Achievement

The biggest news in the achievement area is, of course, the National Assessment of Educational Progress. National Assessment has assumed responsibility for gathering national-level data about educational achievement in 10 learning areas. So far, data have been gathered and reported in seven of these areas.

Predictably, the National Assessment project has drawn fire from numerous sources for numerous reasons. Some people have criticized the learning objectives about which National Assessment seeks to gather data. Still others have criticized the way these objectives were selected and the measures used to gather achievement data. Some fear that the National Assessment data will be used to compare one school with another. Some say National Assessment goes too far, others say not far enough. And there is still a sizeable group of people who look at the very idea of national-level education data as an encroachment upon local control and autonomy smacking of Big Brother.

My purpose here is not to debate the validity of these fears, but rather to highlight National Assessment as a rather bold effort to chart the progress of American education with respect to achievement, and that achievement in this case does not mean earning a degree or diploma, but rather what people actually know or what skills they possess. A recent publication from National Assessment (9, p. 3) sets forth much of the rationale for this national data-gathering effort:

> The fact that disparity in achievement exists is not news. Every teacher knows it; every parent suspects it. Now firm evidence has been collected on a national scale that confirms it and measures the extent of the disparity. But the results. . . do far more than simply point out that differences in educational achievement exist. That fact is only one of innumerable facts one can find, for the results provide a composite portrait of American education . . . . Now, some hundred years after the Office of Education was formed with the charge to do so, it is providing through the National Assessment, information that enables people to act and react with some confidence that their understanding of education in America is based not on conjecture but on solid evidence.

There are other measures used by some as indicators of achievement. Legislators in many states have discovered SAT, ACT, and GRE scores, and regard them as indices of achievement. State testing programs yield data on achievement, but these data are not aggregated to a national level. (Perhaps they ought not to be.)

I have some confidence in the direction set by National Assessment. It may not be comprehensive enough now; it may use some measures that could be refined; it may paint too simplistic a picture of what education is all about. Despite these concerns, I would opt for continued refinement and building of this effort, for I believe the idea does hold potential for successfully monitoring the condition of education with respect to achievement.

Measuring the Impact of Education

Some data are regularly gathered which may indicate various ways in which participation in education impacts on individuals. For example, *The Condition of Education 19*[illegible]11) reveals that the degree of participation in society, as reflected by voting in a major election, varies with level of education. Participation in the labor force is also related to educational attainment, as is earning power. Participation in adult education activities also increases with higher educational attainment.

Numerous studies have been done to examine the impact of schooling on self-image, political inclinations and behaviors, general awareness of social problems, and so on. Once again, how-

ever, we find that this kind of information is basically idiosyncratic in that it is not regularly collected so as to make possible identification of trends with respect to impact variables.

It is, in short, extremely difficult to systematically trace the impact of education. Anyone seeking to do so first encounters the major difficulty of substantiating claims of cause and effect. Our inability to clearly show education (or schooling) as a cause of some defined effect is particularly exasperating to the legislator trying to make choices about allocating scarce resources.

Interpretations made of impact indicators reveal some interesting assumptions. Education is good because it begets more education. Education is good because it enables people to be better employed. Educated people vote more. What we do *not* know, however, is how content people are with their work or whether more educated voters are more enlightened or politically astute voters.

If education becomes increasingly susceptible to public demands for accountability, more indicators of the impact of education will be called for. According to a 1973 Gallup poll, 76 percent of the survey respondents felt that schooling is extremely important to future success. There is still a basic faith in the inherent worth of education. Nonetheless, the value of education may also be a function of its perceived relation to other desirable social benefits.

With respect to resources available for education, existing data are relatively extensive. For example, *The Condition of Education 1975* includes in its chapters on financing elementary, secondary, and postsecondary education 22 charts related to school expenditures. That report also indicates that expenditures by educational institutions and agencies reached an estimated $110.4 billion in 1974-75. In 1973, expenditures in education amounted to 7.6 percent of the Gross National Product, an amount greatly exceeding that of other countries. The Carnegie Commission Final Report (2) describes expenditures, including foregone income, taxes, and implicit rents, of the knowledge sector of American society.

We have, then, regularly collected data that enable us to monitor fluctuations in overall per-pupil expenditures for education as well as changes in personnel costs, operating expenditures for unique programs such as in special education, capital construction costs, and so forth. These statistics allow comparisons of expenditures across states, cities, and school districts. Partly because such statistics *are* available, controversies about the

equities of existing financing structures have increased, resulting in numerous proposed reforms in school finance. While expenditures alone do not tell the whole story of available resources, such statistics do tell much about what is available to do the job.

There are some critical problems involved in the interpretation of indicators of available resources. For example, per-pupil expenditures do not necessarily reflect true differences in effort. Costs for providing equivalent kinds, quality, and quantity of educational services differ in different parts of the country. Furthermore, costs are calculated in different ways by different institutions. The issue of what counts as a cost is particularly acute with respect to technologically based programs that reverse the normal labor-intensive characteristic of most educational efforts. There are as many disagreements about the ingredients of per-pupil expenditures as there are about the ingredients of the Gross National Product.

Furthermore, quality of resources, particularly human resources, is difficult to define. Not all teachers who hold the same degree and receive the same salary necessarily provide equivalent quality of instruction for learners. A poor textbook may cost as much as a good textbook. The climate for learning in a school with high per-pupil costs can be as intellectually or socially stifling as a school with a lower per-pupil cost.

Despite these important problems, much of the basic data on inputs to education is available. Refinements are needed, but a good foundation exists in this area.

Educational Indicators Today and Tomorrow

This has been a very broad look at where we are today with respect to usable educational indicators. My admittedly subjective assessment of the adequacy of our present information base is that we are in relatively good shape with respect to data about access and resources, moderately well off with respect to achievement (particularly in light of both National Assessment and the International Association for the Evaluation of Educational Achievement), and in very poor shape regarding aspirations and impact. My bias about where emphasis on development of indicators should lie is obvious from this assessment.

Some time ago, Senator Walter Mondale introduced a bill in the Congress to establish a Council of Social Advisers that would be responsible for monitoring, on an ongoing basis, specific and actual conditions that affect the social opportunity of people in the United States. The bill also provided for an annual report by the President on the social status of the nation in areas such as education (10).

Such a report would surely require the refinement of a system of educational indicators, focusing perhaps on the areas suggested previously. We would have, in effect, an annual report card of the condition of education. Some have argued that such a report should include an aggregate single index of education status. For example, Ralph Tyler (15) called for a Gross Educational Product. Abbott Ferriss (6) described an Academic Production Index, which would summarize the total production of degrees of all levels. I'm not confident that we'll reach the goal of a *single* index for the condition of education, but we may be able to profit from a *few* indices to tell the story of education.

A coherent information system that allows us to monitor the condition of education over time should provide us with a means of detecting, at various levels, the impact of new programs and interventions in education. If we observe indicators of those things that really matter to us, we should begin to address the issue of whether programmatic efforts result in positive changes in the condition of education. This seems to be the important link between educational indicators and educational policy.

I am somewhat reluctant to suppose that any particular innovation will dramatically affect the condition of education. Education, I believe, is a relatively gentle enterprise whose size and complexity absorb the radical and revolutionary and, for better or worse, neutralizes most things. Change will be incremental; good indicators will pick up these increments. These indicators will not tell us what to do, but will highlight changes.

David Cohen (3) raised some compelling questions about educational accounting: Do we suffer from a short supply of information or from a minimal demand for it? Cohen further wondered about consumer capacity to manage, control, or digest the products of social accounting. Many resources are committed to the technology of gathering and processing information, he argues, but few to its social utilization.

I have touched upon a number of questions that must be addressed in examining the potential for effective utilization of educational indicators. These include:

How can we identify cause-effect, or at least correlational, relationships among variables in education so that we can examine relationships among indicators?

How do we arrive at a useful and valid mix of kinds of educational indicators?

What balance of qualitative and quantitative indicators must be achieved?

How can indicators be "popularized" while retaining an acceptable depth of analysis?

How can the "Big Brother" concern be lessened?

To address these questions requires both an attempt to define more clearly a small set of indices about those things that matter most and to commit resources to serious study of the problems of usage of indicator data. Perhaps the projected leveling of the growth of education affords an opportunity for reflection. With this reflection may come a greater capacity to express and understand the intentions, processes, and outcomes of education.

Finally: Who shall be responsible for developing this system of indicators?

References

1. Campbell, A. & Converse, P. *The human meaning of social change.* New York: Russell Sage Foundation, 1972.

2. Carnegie Commission on Higher Education. *Priorities for action: final report of the Carnegie Commission on Higher Education.* New York: McGraw-Hill, 1973.

3. Cohen, D. Social accounting in education: reflections on supply and demand. *Proceedings of the 1970 Invitational Conference on Testing Problems.* Princeton, N.J.: Educational Testing Service, 1970. Pp. 129-148.

4. Department of Health, Education, and Welfare. *Toward a social report.* Washington, D.C.: Government Printing Office, 1969.

5. Duncan, B. Trends in output and distribution of schooling. In Sheldon and Moore (Eds.), *Indicators of social change: concepts and measurement.* New York: Russell Sage Foundation, 1968. Pp. 601-672.

6. Ferriss, A. *Indicators of trends in American education.* New York: Russell Sage Foundation, 1969.

7. Flanagan, J. Education's contribution to the quality of life of a national sample of 30-year-olds. *Educational Researcher,* 1975, 4, No. 6, 13-16.

8. Glass, G.V. Preface. *Proceedings of the 1970 Invitational Conference on Testing Problems.* Princeton, N.J.: Educational Testing Service, 1970. Pp. iv-vii.

9. Johnson, S.S. *Update on education.* Denver, Colo.: The Education Commission of the States, 1975.

10. Mondale, W.S. Social advisers, social accounting, and the presidency. *Law and Contemporary Problems,* 1970, 35, No. 3, 496-504.

11. National Center for Education Statistics. *The condition of education 1975.* Washington, D.C.: Government Printing Office, 1975.

12. Office of Management and Budget. *Social indicators 1973.* Washington, D.C.: Government Printing Office, 1973.

13. Purves, A.C. Literature education in ten countries: an empirical study. In *International studies in evaluation, Vol. 2.* Stockholm: Almquist and Wiksell, New York: The Halsted Press, 1973.

14. Sheldon, E.B., & Freeman, H.E. Notes on social indicators: promises and potential. *Policy Sciences,* 1970, 1, No. 1, 97-101, 110-111.

15. Stake, Robert. National assessment. *Proceedings of the 1970 Invitational Conference on Testing Problems.* Princeton, N.J.: Educational Testing Service, 1970. Pp. 53-66.

Quality of Life as an Educational Outcome

STEPHEN B. WITHEY
Professor, Department of Psychology and
Program Director, Institute for Social Research
The University of Michigan

The question is: What quality of life accompanies educational accomplishments? In exploring an answer, I will discuss some problems and some findings and make some suggestions. First, I propose to look at some of the problems of assessing the consequences of educational achievement and indicate what some of the consequences are.

Problems and Complexities in Assessing Effects of Education

In a book entitled *A Degree and What Else?* (2), my colleagues and I reviewed studies of the correlates and consequences of educational progress for the Carnegie Commission on Higher Education. Most of my work has focused on that level of educational attainment. We found that some of the correlates of a college education among students and alumni exist because education is not randomly distributed and admission policies and applicants' interests and abilities insure certain specialized characteristics among enrollees and graduates. Throughout most of this century, about half of our high school graduates have gone on to college. This was an exclusive group at one time but in recent years, as high school graduation has become commonplace, college-goers have increasingly become a mixed bag of backgrounds, interests, and abilities, and colleges have changed to serve this growing sector of the population.

There are differences among institutions, so some characteristics will be influenced according to which college you attend

while other characteristics seem to be broadly associated with any college education that includes the extended moratorium and prolonged exposure to the peer society as well as the academic activity of any college experience. Also, the match or interaction between individuals and the college environments they choose is predictive of certain results. How far you go in your education makes enough difference on a few measures so that one can distinguish significant changes with each year or two of education. But there are differences in curriculum, and one can distinguish engineers from English majors on the basis of criteria other than their knowledge of structures and literature. If education is heavily job-oriented, the usual effects are reduced. Whatever the impact of education, its persistence is enhanced by the consistency of activity, conditions, and associates over the ensuing years.

There are a number of characteristics of college graduates that can only be accounted for on the basis of after-college events. Similarly, there are a number of opportunities for, and accomplishments of, graduates that are not simply products of graduation. One consequence of a degree is that those who earn it have an entree into certain occupations that may provide certain income opportunities. These have correlates and consequences of their own but they are not guaranteed by educational achievement. The confirmation or disconfirmation of expectations about post-graduation life and the resulting crystallization of characteristics of the quality of life imagined and achieved by some are part of the prolonged story of educational effects.

Some Educational Effects

The research problem is to sort out those aspects of living that are closely related to educational accomplishment. Entree into certain occupations is possible. Acquisition of certain skills and competence in some areas of knowledge is assumed, and a continuing interest in information and learning is not uncommon. We do know that a desired life style and attendant working conditions are part of the motivation for college attendance. Moreover, college graduates tend to feel better about their work than others and they, more than others, particularly focus on the importance and

meaningfulness of what they do on their jobs. There are also prospects of higher income over a lifetime, and although a differential still exists, it is often exaggerated and, on an average, it is dwindling. We know that values change during college, and although this change seems to be widespread and has been given various titles, the phenomenon seems best explained by the idea of a developed sensitivity to the issues of the period, whatever they were while one was in college. For some, they were the economic issues of depression, for others, they were the issues of war and peace and national systems' rivalries or the issues of civil rights or resources and ecology, and so forth. One may not always end up more liberal or tolerant, though that is a tendency, but one's ideology and values are likely to be molded around the issues of that period of education. College graduates are more thoughtful and deliberate as consumers. They are more involved and informed as citizens and are more likely to participate and assume leadership in organizations and community life.

Not completing college doubles one's chances of not being married, and a number of other socially unstereotypic things seem to characterize the dropout. Having or not having a college degree does not seem to increase or decrease one's immunity to having problems with one's children, but a college degree does seem to make one more introspective about one's problems with children, with oneself, and with others; but along with this sense of doubt or criticism, college graduates report greater marital happiness and sensitivity to the rewards of relations with people. Reduction of obvious prejudice against blacks is principally associated with college training.

These and other factors can be unearthed as consequences of college training with some reservations about what constitutes a college and some accentuation of effects for certain institutions more than others. But these are findings from a rather large group of miscellaneous studies that were not focused on organized approaches to quality of life.

Quality of Life

When one discusses the quality of peoples' lives, one hears many adjectives, and various aspects of life are emphasized. It is clear

that one can characterize quality of life from many perspectives. The notion of quality implies some standard or model or at least a rank order on some significant dimension, or perhaps an optimum or weighted mix of characteristics allowing some substitutions to cater to individual differences. But the question is quickly raised: Whose standards should apply and who should be the judge?

People make such judgments about the quality of their lives all the time, using one or several criteria simultaneously that they think are relevant under the circumstances. They have evaluative feelings about their lives and they make decisions about moving, about jobs and careers, about divorce, degree of endeavor, purchases, voting, and many other subjects. The functioning standards for evaluation are the ones which a person chooses to use, and a common currency for measuring experienced quality of life is the discrepancy between currently perceived conditions and the comparison standards which a person feels are relevant and justified. An obvious implication of this approach is that quality of life can be changed by altering either the conditions or the standards by which they are judged. It is quite possible to characterize the objective conditions of quality of life by such considerations as working conditions, crowdedness of living, physical environment, economic assets, and so forth, but the weight given these factors in the mix of life is essentially a subjective judgment.

My colleague, Frank Andrews, and I (1) studied quality of life from one perspective, and there are admittedly many others. We studied *perceived sense of well-being*, which left to individuals the choice of relevant criteria and the option of using absolute, relative, or specialized factors for their evaluative decisions. Let me explain our model, and then I will provide you with some highlights of the results.

We asked national samples of American adults how they felt about their lives. Answers were obtained on a seven-point scale that ran from "I feel delighted" through "satisfied" and "dissatisfied" to "I feel terrible." The question was asked twice, and individual responses were averaged. Thus, although the scale has only seven points, it was possible for a respondent to pick position 5 and then, when the question was repeated, position 6. Averaging would result in a score of 5.5, which would fall between positions 5 and 6 on the basic scale. Because each question was asked twice and averaged, we ended with seven scale positions and six posi-

Table 1

How do you feel about life as a whole?

	(7) De-lighted		(6) Pleased		(5) Mostly Satis-fied		(4) Mixed		(3) Mostly Dissat-isfied		(2) Unhappy		(1) Terrible	
May '72	7%	9%	24%	22%	23%	7%	5%	2%	1%	*	*	*	*	
Nov. '72	6%	7%	23%	18%	26%	8%	6%	2%	1%	1%	1%	*	*	
Apr. '73	7%	9%	24%	19%	22%	8%	6%	1%	1%	1%	1%	*	*	

*Less than 0.5%

tions between scale anchor points. Distributions obtained on three studies are given in Table 1. Although satisfaction seems widespread, one should keep in mind that 1 percent may represent about 1,400,000 people.

In case this picture looks deceptively rosy, let me add that most people see flaws in the quality of their lives, and about two-thirds of the population want to change some aspects of their lives. Also, it is clear from some of our measures that feeling dissatisfied with part of one's human condition does not have much effect on overall evaluations. Ratio measures indicate that people are somewhat accepting and tolerant of a few flaws and will still report that they are "mostly satisfied" with their lives. The mid-point of the scale is clearly not the median of the distribution, and we regard even expressions of "delighted" as including a fair number of discontents. Apparently it is true that, in our culture at least, evaluating "life as a whole" does result in some tolerance and minimizing of life's dissatisfactions along with some optimism that things will work out.

We theorized that this overall evaluation of the quality of life was a summation of, and could be predicted by, evaluations of less global aspects of life. Surely how one feels about life as a whole is a product of how one feels about one's family, job, home, neighborhood, community, and so forth. These separate areas might be termed *domains* of life. But when one judges the quality of, say, one's family life against some image or reference, the latter contains specific *values* and criteria. Family life, for instance, might be judged by love, cooperation, fun, loyalty, respect, novelty, and so on. And these same values might be found in some other

domains too—fun with friends, respect on one's job, cooperation in one's neighborhood.

Imagine a matrix in which the rows are domains of life and the columns are the value criteria by which quality is judged. An individual evaluates the overall quality of a domain by summing across relevant values. Similarly, people judge the overall satisfaction associated with a value by adding value-satisfactions across relevant domains. For instance, in evaluating one's neighborhood, judgments are made on criteria of attractiveness, safety, traffic levels, friendliness, convenience, and so on. Or evaluations of one's security might well involve judgments about safety in the neighborhood, on the job, and while traveling; economic stability; national security; and so on. One test of such a model is to compare the predictability of global, life-as-a-whole evaluations based on domain satisfactions with their predictability based on value satisfactions. It works about equally well either way. About 60 percent of the variance in global evaluations of quality of life among the population is accounted for either by how people evaluate domains or how they evaluate the satisfaction of their value criteria.

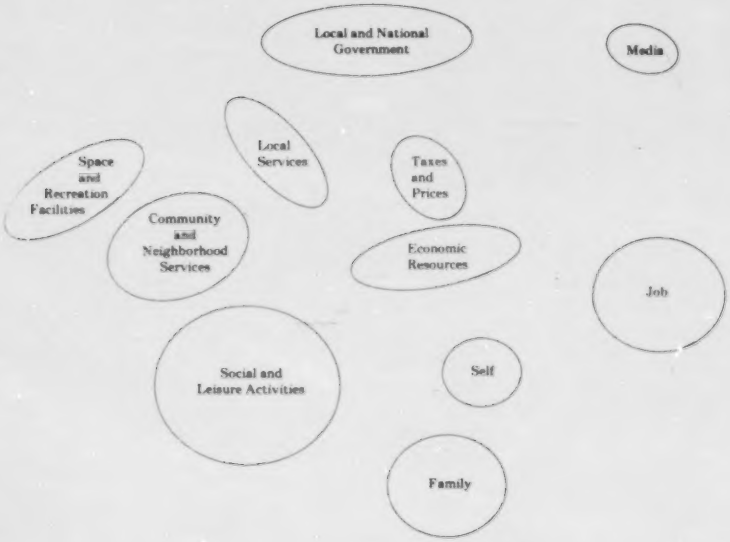
The Structure of Life Evaluations

These analyses tell us something about how people judge the quality of their lives, but it would be revealing to know more about the structure of these judgments. Let us look closer. One can clearly divide life into a large number of domains defined by places, people, and activities. Maybe some of them cluster together so that feelings about one item in the cluster are close to predictable from evaluations of other items. These clusters, which tend to share evaluative feelings, could then be regarded as significant domains of life and they would reflect how most people break up their world for judgmental purposes. Using Smallest Space Analysis and a cluster analysis, such clusters can be identified. They hold for the population as a whole as well as for various subgroups such as old and young, rich and poor, and educated and uneducated, though there is some variation in the independence of clusters from each other among different social subgroups.

In all, we tried about 123 items—some domain items and some

value items—on various national studies, but not all of them on any one national sample (although we did try all items on about 220 subjects in one city). On all measures, there were some occupants at every point on our seven-point scale. The results of one such analysis using 62 domains yielded the 11 significant clusters using convex polygon criteria summarized in Figure 1. This structural analysis has been replicated. Question items that fall within the "Family" cluster are: How do you feel about your children, your wife or husband, your marriage, things you do with the family, your family responsibilities and, marginally, your relatives? Items called *Local Services* include medical, transportation, home repairs, and commercial services in stores; and *Economic Resources* include job pay, family income, and standard of living. Actually, these clusters are better handled in three dimensions, but two do not work badly and are more easily read. It is not easy to label dimensions on these maps, but one is clearly a dimension with private close-to-home domains at one end (bottom

Figure 1
Map of Significant Clusters of 62 Domain Items



of Figure 1) and shared public domains at the other (top of Figure 1). Average satisfactions with domains fall in the same order with the highest or most satisfying national averages being reported for family and friends and the lowest or most dissatisfied national averages being reported for governmental activities, general economic conditions, and the like.

A similarly plotted map of value criteria shows a different structure, one in which there is a core set of values with a set of Saturn-like rings and then a more distant set of satellite clusters. There seems to be less discrete structure for the value items, which suggests that it is indeed hard to be pleased with the amount of fun in one's life without at the same time being fairly satisfied with one's security, the amount of respect one gets, and so forth.

We have already reported that evaluations of these areas of concern predict most of the predictable variance in global life evaluation, but can we do as well with a shortened list? We find that we can. Using a Multiple Classification Analysis (though it works as well with a regular multiple-regression analysis) and selecting those items with the highest betas, we find that we can predict between 50 and 60 percent of the variance in global life evaluation with only 12 items and just under 50 percent with only 6 items. Will any small set of items do as well? No. The items that work best as a set of predictors require some representation from each of the clusters that make up the structure of domains or of values as we have reported them. The domains that have the most influential beta weights, for example, are those that fall in the lower half of the cluster map, but some improvement in prediction is added by including items on government, consumer items, and so on. These analyses add strength to the argument that these clusters have some real significance in peoples' appraisals of the quality of their lives.

After rather extensive analysis, we came to the conclusion that a weighted *additive* combination of evaluative responses is adequate to capture virtually all the predictive power present in these clusters. We had thought there might be substantial interactions in the data, but so far none of any significance has been found. We may conclude, therefore, that most, but not all, of what constitutes quality of life is made up of what might be termed that part of life that is close to home, involving such areas of concern as one's self-efficacy, family, money, fun, housing, and neighborhood.

Prediction from Education and Other Demographic Variables

How well can one predict global quality of life from such variables as sex, income, education, socioeconomic status, age, and life-cycle? The answer is that only 5 percent of the variance in life-as-a-whole measures is predicted by all these considerations combined. In this poor showing, the best predictors are family-life cycle—partly because individuals whose family life has crumbled report a rather miserable quality of life—and the level of family income. Education, as a variable, comes up with a beta of only .06. This does not mean that there are no significant differences by social or demographic subgroups in domain and value satisfaction. Although we find significant differences in global quality of life across socioeconomic status, the differences are not large enough to account for much of the variance. Education by itself, instead of as a component of socioeconomic status, works even more poorly. We do find that the same predictive pattern from domains or values to global life evaluation essentially holds for some 20 or so social subgroups we have analyzed. So the structure we have reported is not just the result of averaging across the total population.

There is a slight tendency for those with a college education (but not as true for those with postgraduate degrees) to report better assessments than others do of their marriages, their spouses, their children, young people in general, their own health, and some of the aspects of their jobs if they are working. But as we have already noted from other data, although happiness with marriage increases with level of education, feelings of inadequacy as a marriage partner also increase with education, and similarly, feelings of criticism of oneself as a parent also increase.

The only other concerns that show decreases in satisfaction along with rising levels of education are: time to do things you want to do, the informational mass media, and the condition of the natural environment.

If one combines education and income in associated rising steps, excluding those individuals who have high incomes with low education and those who have college education but relatively low incomes, many of the differences just listed are accentuated, and satisfactions with economic conditions are added as aspects of life that are increasingly satisfying as one moves up the socio-

economic ladder in both education and income. For instance, associated with the progressive steps of joint educational-economic attainment, the group means on life-as-a-whole measures are (on our 7-to-1 scale shown in Table 1) 5.0 (low income, under $4,000; low education, less than high school), and by steadily rising steps go up to 5.7 (moderate income, over $12,500; high education, college). But higher education with low income drops the group's average back to 5.2, not much above the poor and uneducated. High income with low education produces only 5.4, still not up to the level of joint achievement in education and income.

The Education Variable

Apparently educational attainment does not necessarily improve the quality of one's life as perceived and evaluated subjectively. Let me propose five factors that may contribute to such a finding:

1. What education may accomplish is a raising or at least a changing of standards and criteria for judgment. Higher aspirations and higher demands on life, if they accompany education, could lead to a constant discrepancy between standards and perceptions of life as standards and conditions both improve.

2. Another thing that may accompany educational accomplishment is greater honesty in appraisals with a recognition of problems and failures along with the greater sensitivity to potentials and opportunities. Clearly, in some areas of life, it seems that education leads to an increase in mixed appraisals and more consideration and tolerance for such a mixed picture.

3. Education does not always lead to accompanying improvements in other components of social status. When the characteristics that are usually associated with status, such as better occupations with higher income following educational achievement, are all achieved, social benefits are more probable than when the expected pattern of growth of social status is disrupted. It may be that any muddying of the patterns of social achievement is discouraging and disturbing and accounts for some of the failures of educational achievement taken alone to indicate the hoped-for improvements in quality of life.

4. One of the problems in assessing the consequences of education is that education seldom seems to operate the way a good continuum should. The characteristics of those who do not achieve the programmed levels such as a high school diploma or a baccalaureate degree (dropouts) do not always place them neatly between their neighboring groups on the educational progress ladder. And in some ways, those who stay on for postgraduate degrees end up no better off than those who stopped with a bachelor's degree. It may well be that education looked at as years of training has sharp differences of meaning at different points on the continuum, and these should be considered in insightful studies of the impact of education.
5. As we reported early in this paper, higher education has become a rather heterogeneous mixture of educational offerings serving a broadening range of students with increasingly variant interests, skills, and aspirations. It would be surprising if this increasing variety led to similar effects and it is indeed surprising that in some aspects it seems to do just that. Nevertheless, it may be that one should be more discerning and sensitive in separating types of educational experience for evaluative studies of consequences and effects.

Suggestions

What, then, might be proposed for research on indicators to be tracked over time?

1. We have developed about a dozen usable and tested indicators of feelings about domain and value clusters that seem to adequately cover perceptions of well-being, and these can be monitored.
2. Instead of lumping levels of educational attainment together, we might look at people who achieve higher levels of quality of life and look at indicators of their educational experience to find out what aspects of education have a payoff in quality of life.
3. Instead of assessing only cognitive and affective appraisals of life conditions, we might also look at the standards and comparison referents that people use for their judgments.

4. Another consideration is associated with the map of domain concerns. On that map, you will remember, most people's satisfactions were reported in the lower third of the map—areas of life with which people had direct contact, experience, and interest. It may be that education enlarges the world of interest and experience and increases the salience of those areas of life that most people, of whatever education, tend to find most dissatisfying in present times.

5. Since judgments of worth and quality applied to components of life experience are of uncertain significance, we might look at the implications of experiencing various levels of quality of life in the behaviors, activities, and endeavors of people. We know that various factors such as optimism, reference groups, and personal defenses creep into the evaluation process and, although we may all aspire to higher levels of life satisfaction, we are not clear on the implications of occupying them.

6. In addition to monitoring various measures and components of quality of life, we need to look at how the parts fit into a larger picture. It may well be that the interrelationships of parts of life may undergo more significant flux and change than would be apparent from the figures on isolated areas of life tracked by themselves.

References

1. Andrews, F.M., & Withey, S.B. Developing measures of perceived life quality: results from several national surveys. *Social Indicators Research*, 1974, 1, 1-26.

2. Withey, S.B. *A degree and what else?: correlates and consequences of a college education.* New York: McGraw-Hill, 1971.

Measurement and Efficiency in Education*

MANCUR OLSON
Professor of Economics
University of Maryland

I should like to begin with a striking statement made by a distinguished social scientist who often studies education. The statement was made in a small, informal seminar in which there may have been an implicit assumption that any possibly embarrassing comments would not publicly be attributed to the participant who made them, so I won't identify the source. The statement was that it was "surprising" (note the word "surprising") that in many parts of the country many of the best superintendents of schools had started their careers as coaches of high school athletic teams. In using the word "surprising" I don't think my friend intended, and I certainly don't intend, to belittle coaches of athletic teams as a group. I don't want to be like the nineteenth century writer who described a cavalry officer he particularly disliked as being so stupid his colleagues noticed it.

My purpose is rather to talk about the incentives, information flows, and selection procedures in public schools and in government generally. I hope to be able to relate the comment about the "surprising" success and achievement of coaches in managing teachers of subjects very different from their own to these incentives, information flows, and selection procedures, which are, in turn, tied to some of the crucial problems of public education today. They are tied to the fact that seniority determines pay levels, and even sometimes responsibility levels, in many public school systems in the country; to the fact that there is so much red

*The author thanks the National Science Foundation and Resources for the Future for supporting his research on aspects of this topic.

tape in our school systems; to the fact that throughout the country, parents, teachers, students, and observers alike have the gut feeling that somehow we could be educating our children a great deal better. And this sense that we could be doing much better is not peculiar to this country or to the last few years; it seems to be a rather general reaction to the public schools that are almost everywhere the dominant element in the educational system.

I

To get the particular perspective on these problems that I believe is needed, we must, it seems to me, start first with two concepts. The first of these is peculiar to my discipline of economics. It is the concept of the "public" or "collective" good and the distinction between it and the familiar notion of the "private" or "marketable" good. The "goods," or valued things that we get from the government such as law and order, pollution control, and defense, are collective goods and inherently different from the private goods we buy in the stores. The single most important difference is that the public good is such that if it goes to anyone in some group, it goes to everyone in that group. Thus, if the air over New York City is cleaned up, it is cleaned up for everyone in New York (or in any event, for some large group of people such as everyone in the relevant airshed). Similarly, if the East and West Coasts are defended against amphibious attack, the Midwest is inevitably also defended. Again, if the police and courts arrest a burglar caught breaking into my house, my neighbors are also made more secure by his incarceration.

If the group that benefits from a public good is very large, business operating in the marketplace will not be able to provide it. Public goods usually have to be provided by governments—that is to say, you have to use compulsory methods like taxation to get the needed revenue. I wish we had time here to go into the different types of public goods and their importance for public policy, for this is one of the most important and interesting topics in the field of economics today.

In the case of education, the situation is made a bit difficult or complicated because education is, in part, a public good and, in

part, a private good. One aspect of education that's definitely public is that which is designed to make the democratic political system function better. It is quite clear that if my children are getting a lot of civics and the civics makes them better participants in the democratic process, then there is ultimately a gain for people other than my children. That is to say, there is a benefit, like the benefit from pollution control, that inevitably reaches some large group—in this case, everyone in a particular polity or a political system. At the same time, though, we know that by no means all of education is like this. Education also passes on marketable skills—frequently skills whose benefits are captured entirely by the person who possesses them in the form of wages from the employer who hires him. In this case, of course, education, even if provided by governments, is a private or marketable good. I hope it will soon be clear why I think this distinction between private and public goods illuminates the issues of educational testing and assessment that this conference is about.

But before we go on, we must talk about the second concept, or tool of thought, that's needed at the moment and that is the particular definition of social indicators that is required, especially in this context. As you were told by Eleanor Sheldon earlier this morning, there is not yet consensus about what definition of social indicators is most useful. But at least for the purposes of the present argument, it is clear that we must define a social indicator as some measure of social performance—that is, as a measure of outputs or variables of direct normative or welfare interest. Thus, a measure of how much children have learned is a social indicator: It tells us something about our society's future well-being. This definition of social indicators is the one I introduced when writing the opening chapter of *Toward A Social Report.** (Since comments earlier this morning have attributed this report exclusively to me, I want to emphasize the invaluable help of my staff and consultants and the decisive administrative support for the project from Wilbur Cohen.)

Now, where do we go with these two concepts, with the idea of public goods on the one hand and this particular definition of social indicators on the other? These two concepts will give us an

*U.S. Department of Health, Education, and Welfare. *Toward a social report.* Washington, D.C.: U.S. Government Printing Office, 1969.

insight into the efficiency or effectiveness of schools and other agencies of government. If it's the case, as I believe it is, that we want more of the desired outputs of education, if possible with less resources, then we want efficiency. To analyze this, we must look into the relationship between the inputs we would like to be able to use less of and the outputs or results we would like to increase or improve.

II

A specification of the maximum amount of the desired output that can be obtained from a specified amount of resources, given the current state of knowledge about how this production can best take place, is called a "production function." or sometimes in contexts in which public goods or social purposes are dominant, a "social production function." To analyze whether an activity is efficient (or as effective as it should be, given the resources available to it) we must know what output it has and, in addition, ask what results it obtains in comparison with the best results that are possible with the current state of productive knowledge, for if the social production function itself is now known—if we don't know what would constitute top performance—we cannot properly assess whether or how far current performance falls short of what should be expected.

My most basic argument in this talk will be that where public goods are concerned, it is unusually difficult either to measure outputs or to estimate production functions and therefore exceptionally difficult to assess effectiveness and to operate efficiently. This argument can best be understood if we think initially about unambiguous public goods like pollution abatement, national defense, or law and order. After we have done this, we can more readily adapt the argument to our special concern with education and testing.

One reason why it is typically harder to assess the efficiency with which public goods are or could be produced is that measuring the quantity or volume of what is being produced is usually far more difficult. This greater difficulty grows out of the defining characteristic of public goods mentioned earlier—the fact that the good, if it goes to anyone, goes to everyone in a group because of

its inherent indivisibility. This is evident in a particularly striking way when we compare a firm selling a tangible private good with a government department providing a collective good. The firm dealing in a private good can *count* the number of units of output it produces or sells, whether they be automobiles, television sets, or loaves of bread; it knows its quantity of output and would be able to get information about its "technical efficiency"* even before it had sold its output. By contrast, no government or other producer of a collective good can count or directly measure the units of output it produces and thereby check its technical efficiency. To be sure, an army can count its tanks, a police department can count the miles its patrol cars have been driven, and a public health service can measure the insecticide it has spread. But these are "intermediate goods" or goods used to produce other goods. Most of these and other inputs used in producing a collective good are private goods that can be counted, but the production function cannot be known unless the final outputs as well as the inputs are known. What is needed is not simply the number of tanks but also the reduction in the probability of attack or defeat; not simply the miles patrolled but also the reduction in crime; not simply the insecticide spread but also the number of infections prevented. The latter outputs cannot be directly counted or measured.

To be sure, the instances just cited are special cases that are obviously favorable to the argument being made, because tangible consumer goods are compared with public goods, all of which are really services. Ultimately, all goods are valued for the services they provide—the automobile is wanted because it can provide transportation service, the television set because it provides entertainment service, and so on. All services are intangible and, thus, not generally capable of being measured in simple physical units. This seems to suggest that it is equally difficult to determine the quantity of output of private and of public services. The truth is, however, that it is systematically more difficult to determine the output of public goods than of private goods, even when the latter are plainly and immediately services (as opposed to goods that the consumer uses ultimately to obtain services).

*An undertaking is "technically efficient" if it cannot, with the best available technology and the existing resources, produce more of one good without less of another.

This is because the producer of the private service can get some insight into the quantity of output he is producing by analyzing his total revenue. This is, of course, the product of the quantity of service he sells times the price he receives per unit of service and can vary because of autonomous changes in demand curves as well as changes in the quantity of services the producer provides. But if the producer can either independently identify the changes in demand, or know that they are randomly determined, he can gain insight into changes in the quantity of output he is producing from the variations in his total revenue. The agency producing a public good has no such opportunity.

The community may gain knowledge of the output of a collective good by experiment. It may gather some evidence on the effectiveness of a Maginot Line by relying on one, or insight into the effect on crime of doubling the police force by trying it, or information about the effectiveness of advertisements for safe driving by examining the accident rate after a publicity campaign. Societies may also learn something from one another; if the breathalyzer cuts down accidents from drunken driving in Britain, it may do so elsewhere as well.

III

The special difficulty with public goods—and the second reason why it is harder to determine the efficiency with which public goods are produced—is that any experiment affects everyone in the area or group that receives the public good. Since communities and often whole societies and (on rare occasions) even the whole world must experiment to get information about what happens when a new method of production or allocation of resources is tried, there cannot be as many experiments with public as with private goods or as much useful experience gained in a given amount of time. There are, for example, fewer experiments on the effect of air pollution on health than on the effect of cigarette smoking on health. This is probably one of the reasons why the effect of cigarette smoking on certain health problems is well known whereas the effect on health of typical levels of air pollution remains relatively uncertain.

Consider also the desire to protect the country against attack or

defeat by a foreign power. Note that this not only raises questions about what kind of weapons it is best to have, how large or small the defense budget should be, what generals and what Secretary of Defense ought to be in charge, but also whether maybe in some cases, you wouldn't do better to have a soft-spoken foreign policy that makes it clear that we will sometimes compromise differences with other countries rather than spend more on our military strength.

How could we get empirical information on the important question of how we should make ourselves secure as a nation? How could we estimate the social production function for national security? Surely we are not going to experiment by first arming to the teeth to see what happens and then disarming altogether in order to get a little information about the effects of that type of policy. Obviously these could be very costly experiments. The point is that the impact of our foreign policy and our defense policy is a public good that reaches so many millions of people that changes in policy designed to tell us more about the cause-and-effect relationships involved in producing military power or international stability can be incredibly costly. Because the number of major nations in the world is limited, we also learn only a limited amount from the experiences, or "natural experiments," that take place over time. This is an area of *ideological* debate, where even the smartest and best informed people don't often come to agreement even over long periods of time.

Similarly, there is continuing disagreement about how much police protection there should be, what kind is best, and so on. Again, the explanation is that whole cities tend to benefit from police protection.

If, by contrast, we are talking about a private good, like bread, and asking how we could learn how to produce more wheat for less money, it is clear that on some tiny plots of land costing very little indeed we can test out the yields of alternative varieties of wheat, the output gained from marginal amounts of fertilizer, and so on. Because solid empirical information can be got rather cheaply, there are normally no ideological debates about the production functions for private goods. Debates are settled empirically because the empirical information needed to settle them can be obtained and this information is normally attainable chiefly because the goods are private goods.

IV

When we turn from those activities that produce goods that are unambiguously public to the case of education, the situation becomes more complicated. If we didn't care about the effect of schooling on the functioning of our democratic political system but only about the extent to which schools prepare people for the job market, and if the schools were independent schools that financed themselves by competing for tuition-paying students, then the argument that has been made about public goods would have no application to education.* Of course, the statement I have just made does not exactly describe the world we live in. In our world, education is usually provided on a basis that what one kid gets a whole group of kids will be required to receive. The world doesn't have to be that way, but that is the way our world is, and to that extent, the argument I have made in this talk is applicable. In addition, most of us are anxious to preserve and improve the democratic process, and to the degree that we have this goal, it is not possible to make education a private good, however much we would change the educational system. Thus, the argument put forth in this talk clearly applies to education, but not to all parts or aspects of the educational system.

If the argument here is true, it leads to some important implications or predictions about the characteristics of government agencies producing public goods and also to a great extent about the characteristics of educational institutions as well. The first implication or prediction is that it would be frightfully hard to get a high level of efficiency in government agencies producing only pure public goods and even rather hard to get maximum effectiveness in most school systems. This, in turn, would lead us to expect that citizens repeatedly will have the gut feeling that somehow, with the resources and technology that are available, it should be possible to do a lot better. Without being able to prove it, many citizens are likely to think their governments and schools could very likely do much better, and they are right.

*It is *not* the absence of government that makes a good a private good; the government could provide aid to students from low-income families, or to all students, but the public good argument set out here would still not apply if the facts are as described in the sentence in the text to which this footnote is appended.

Another implication is that agencies producing unambiguously public goods—and to some extent, educational systems as well—can have little incentive to be efficient. The intentions of the higher officials and of those who designed the institutional arrangements may have been to reward efficiency, but it is not possible to provide an incentive to be efficient to an agency or agency head if there is no way of knowing whether the agency at issue has been efficient or not. The agency and agency head will normally have an incentive to *seem* to be efficient, but this is as likely to lead to expenditures on public relations as to more effective or economical policies.

A further implication of the argument is that effectiveness will not usually be the dominant determinant of what personnel, procedures, or policies are chosen in institutions producing public goods. If the true output of a public agency is not known, neither can the contribution to output of any particular employee, procedure, or other input be known. If decisions are to be made about which employee should be promoted, they have to be made in large part, if not entirely, on some basis other than a record of efficiency. Because educational qualifications, expository skill, and a style that fits in with the cultural, political, and administrative ethos of the time often play a large role in personnel recruitment and promotion in public agencies, the "merit system" has not usually troubled most intellectuals. But to demand civil servants with greater educational qualifications is to demand more expensive inputs, and such qualifications may not lead toward a more efficient provision of collective goods. Similarly, decisions about other inputs or about alternative methods or technologies for providing collective goods must be made on some basis that, however likely it is to accord with currently fashionable doctrine, bears no necessary relationship to the cost-effectiveness of the different alternatives.

In the case of the military, especially in peacetime, this comes out with particular clarity. Have you ever asked why military officers are supposed to have especially well-shined shoes? Obviously, you can fight just as well in unpolished shoes, and in the thick of battle, they may be allowed. The answer, if my conception is even half-right, must have something to do with shoeshining being a proxy measure of the devotion to duty and effectiveness of the officers and sometimes when no fighting is under way, pos-

sibly as good as other available proxies. This is, of course, somewhat of a frivolous example but one which may, nonetheless, indicate the essence of the problem that we are dealing with.

V

In government agencies producing public goods, we find civil service systems, merit systems, an unusual degree of credentialism, lots of emphasis on following procedures in staying out of trouble with superiors, and so on. And the reason we find these things is not because people have perversely wanted to use the wrong measures to assess performance; it is simply that it is difficult to get the right measures. I sometimes speak of the "4-S criteria"—schooling, shoeshining, seniority, and sycophancy—for retention and promotion in agencies producing public goods.

Now this, of course, overstates things, especially in education. In the case at least of private schools, there is some measure of effectiveness of a particular headmaster, for example, in whether or not he is able to make the school attract enough tuition to stay in business. So, of course, I am looking at a polar case for the sake of brevity and clarity, but the world is far more complicated than that.

Let us also look at the contrast between the State Department and the civilians in the upper reaches of the Defense Department in the U.S. Government, particularly in the fifties and sixties. It is well known that, over most of our recent history (except possibly during and since the Vietnam war), the State Department has been able to attract college graduates with the highest grades and the best recommendations. And my observation is that most foreign service officers are, in fact, quite talented. But have you noticed the poor reputation of the State Department as a bureaucracy, as an *organization?* I am not talking about Secretary Kissinger personally or about his predecessors; I'm talking about the State Department as a collectivity. We know that again and again, Presidents have relied on a national security office in the White House rather than the State Department. And it has been a common assumption in Washington at various times in the recent past that at interagency meetings and in other contexts, the civilians from the Defense Department seem to do on the whole

rather better than the bureaucrats in the State Department. Of course, these perceptions may have been wrong, and by my own argument they must be viewed with caution. But let's ask how we get civilians for upper-level civilian jobs in the Defense Department. Certainly people don't go into the Defense Department civilian sector at the GS-7 level or the GS-9 level for the prestige involved. There are problems of recruiting really able people at that level in the Defense Department. Most of the higher ranking civilians in the Defense Department are businessmen who made money, politicians who picked winners, lawyers who won cases, professors who wrote famous books—people who have succeeded in some line of endeavor *other* than the Defense Department.

Now, of course, it doesn't follow that success in one thing will necessarily assure proficiency at something else. But note that when people come into a bureaucracy later in life, having been doing something else, they will be different in two ways from those who have been there all along, if it's an agency that produces public goods. First, they are sure that things *can* be done differently because outside they *are* done differently. They are likely to be less conservative in a bureaucratic sense. Second, they may have come from an area where keeping score on what is being accomplished is a little easier than it is in the area of public goods. To the extent there is any correlation between achievement in the area where someone is keeping score and efficiency in producing the public good at issue, this is a significant matter.

Thus, it is by no means obvious to me that policemen ought always to be recruited at age 20 or so and police chiefs chosen only from people who spent a whole life in a given police department. You get very particular kinds of conservatism and ideology from that. The military, fire departments, and foreign services are much like police departments in this respect, and in some schools, somewhat similar practices prevail. In other words, there is a need to get fresh thinking from the outside, especially from outside areas where someone is keeping score.

VI

Now we come back to our old friend, the successful coach who later becomes a successful superintendent. Surely physical educa-

tion is not the most central or typical subject in a school curriculum. But the coach guides a team that plays other teams. Whenever the teams compete, there is someone keeping score.

To be sure, this point about coaches is not the general one that I am trying to emphasize; it is only an instructive special case. It should, however, encourage us to search for more general and apposite methods of keeping score, which brings us back to our second tool of thought—the concept of the social indicator.

Consider what happens in the absence of the testing and measurement of the sort many of you do, and of the sort the organization that sponsors this meeting does—that is, measurement and testing designed to assess the performance of schools, methods, and teachers, not simply the performance of individual students. In other words, in the absence of the social indicators that testing provides and in the absence also of competition from a really large private school system, there is no way to know how well a school is doing. There will be the familiar difficulty that we won't know how to allocate resources efficiently and, thus, will pass up opportunities to get more for less. Moreover, there will be the further difficulty that there is likely to be what might best be called "organizational arthritis"—lots of seniority determining who gets paid how much, lots of credentials determining who gets promoted and who does not, lots of red tape, and lots of concern about the particular events that happen to be newsworthy rather than about some others that despite their importance are not. Organizations that produce, even in significant part, public goods have, because of the special difficulty of getting information, a special tendency to contract institutional arthritis.

All that I have said about tests as social indicators must be understood in the context of "multiple causation." As everyone now knows, there are many things *besides* the schools that affect how much is learned. Home environment, community structure, genetic endowments, and television are obvious examples. It follows that the degree of success that will be obtained from using testing to improve resource allocation and combat institutional arthritis will be very limited. There is, for example, no way you can adequately test to see what it is about an educational system that will make the people that come out of it have a successful democracy. You can test for knowledge of what is given in a civics book and that may or may not be correlated with a successful

democracy, but it is clearly not the same thing as a successful democracy. And if a given country fails in democracy, that doesn't necessarily mean that its educational system is the cause; we can't prove, for example, that because Hitler rose to power in Germany, German schools were uniquely bad, or even bad at all.

VII

Thus, for fundamental conceptual reasons where public goods are involved, we must be satisfied with the most scattered and miniscule successes at best. But this infrequent and tiny success, it seems to me, can be worth a very great deal. Test results as social indicators are rather like cortisone for the person with arthritis: They cannot possibly cure the disease and often even have adverse side effects. Yet they may, in a small way, make it possible to use our educational muscle a little more effectively and at the least encourage a bit of movement.

Luncheon Address

Educational Indicators and Social Policy

WILBUR J. COHEN
Dean, School of Education
The University of Michigan

I don't know that I'm going to agree with what I m going to say today. I feel somewhat like Senator Everett Dirksen must have felt one time when he got up to speak to a group. He was handed his speech by one of his assistants and he read the first page with great enthusiasm and the second with equal enthusiasm, and went on to page 3 and 4. When he had read part of page 5, he stopped and said, "Hell, I don't agree with this at all!"

Having had speeches written for me in the past, I've felt like that a number of times. Now that I am out of office and have to write my own, I am more concerned with the validity of what I say, particularly with such a distinguished audience as this one.

In my diversified career, I have had two experiences that influenced my attitude about transforming data into effective use for policy decisions. The first occurred in 1939 when I was a young research assistant responsible for supplying technical information to people at the legislative level of decision making. I had gone to a meeting of the so-called Conference Committee between the House and the Senate. The chairman of the Conference Committee at that time was Senator Tom Connally of Texas, one of the flamboyant old-type senators with a bow tie and a great deal of rhetoric who was very self-assured about his role as a senator. The issue under discussion concerned an amendment that Senator Byrnes of South Carolina, later a presidential assistant and Secretary of State, had offered to change the welfare system. At a certain point in the discussion, I was asked to estimate the cost of the Byrnes Amendment. As a young civil servant filled with the pride of technical competence, I answered that the amendment would cost

somewhere between 80 and 120 million dollars a year, a monumental amount then. Senator Connally looked at me and said, "Don't you know?"

I said, "Well, Senator, it depends on the assumptions that you make with regard to what the states will do."

"Mr. Cohen, you know we in the Senate give a lot of money to the Executive Branch for research and study and measurement of all these problems, and here it is four-thirty, quarter to five in the afternoon, we ask you a simple question—how much does it cost—and you start telling me it costs A, and it could cost B, and up to C. When we ask you a question, we would like to have a frank, simple, honest answer."

"Well, Senator," I replied, "I tried to give it to you, and if you'll give me a chance to discuss the different assumptions, I'll tell you."

"Well," he said, "it's too late now. Let's go home and you come back, Mr. Cohen, in the morning with a definite answer, will you please."

Well, I went back to my office and finished doing the routine things and went home and then found that something was wrong with the children. Ten o'clock in the morning came without my having done anything about the Senator's order, and when we came back to the Conference Committee, Senator Connally said, "Do you have an answer, Mr. Cohen?"

I said, "Yes, sir."

He said, "What is it?"

I said, "A hundred million dollars."

He said, "That's fine, that's the kind of real information we like to have."

I learned many things from that experience, one of which is to give congressional committees simple answers to simple questions. The absorption capacity of senators and congressmen is very limited, and if you try to give them too much statistical information all at one time, it casts serious doubts on your competence as a research person.

My second experience, which I think also had a tremendous effect upon me, happened somewhat later in the Hyde Park Franklin D. Roosevelt Library. I was doing research on some of the origins of the New Deal and the role of some of the outstanding leaders in formulating policy during the twenties that had an

impact on developing the New Deal. I've always been deeply interested in the origins of policy and development: How did they occur and sustain themselves, who brought them up, and who were the godparents and the mothers and the fathers of those ideas? While at the library, I came across the second draft of Franklin D. Roosevelt's historic 1937 second inaugural address. I was absolutely fascinated by it because it had several people's handwriting on it including his own (he had very distinctive handwriting, and without being a handwriting expert, I could tell which writing was his). Somebody had presented him the draft, which said, "I see twenty-five million people earning $600 a year," and he had crossed $600 out and put in a weekly amount. He didn't like using annual earnings, I would assume, because he said that nobody knows how much they make a year, they know how much they make a *week*. He didn't ask anybody whether the annual earnings should be divided by 42 or 50 or 52; he just divided the annual figure and came out with $20 a week. The draft stated that there were millions of people with such-and-such income, and without benefit of any research, he decided to add that this was six million families. The draft kept referring to "millions of people" doing this and "millions of people" doing that, and I think he was disturbed that some people didn't really know what millions of people and millions of dollars really meant or what its significance was. It was then that he wrote near the bottom of the page, "I see one-third of a nation ill-housed, ill-clad, ill-nourished." I have often wondered what went through his mind as he wrote. He could have written, "I see 34.7 percent of the nation living in sub-standard housing, as defined by the Commerce Department, and 21.5 percent without electricity, according to the standards of the electrical industry, and 32.3 percent without adequate clothing for at least one member of the family, and somewhere between 31 and 41 percent of the population malnourished or undernourished, depending on the particular criteria of the National Nutrition Council that you use."

I think Franklin D. Roosevelt created one of the most powerful and significant social policy indicators of his generation when he created that statement. As a social policy, it influenced my whole professional life. I have spent 40 years trying to deal with that single indicator, that single statistic, and do something about it.

I cannot throw much light on problems of educational indica-

tors because during my work as Secretary of HEW, I came to the conclusion that educational indicators are so much more difficult than health and welfare indicators that I tended to concentrate on the latter. So today I want to take a health indicator for you and show how it can be converted into social policy. I will attempt to illustrate how fundamental your work on social indicators is in helping to formulate new social goals and social change.

Infant Mortality as a Social Indicator

Infant mortality is a social indicator whose significance in terms of social policy is both widely recognized and widely disputed. Like an economic indicator, it comes out every month. It is produced by the National Center for Vital Statistics so that it is in the public domain and is available nationally on a state basis and periodically by Census tracts. It is a well-known indicator; it's available; and for many people—at least nonmedical people—it represents a very significant indicator bearing on the quality of life: Most people assume that when infant mortality declines, that is a social good and when it increases, that's bad. So infant mortality represents a well-known, significant social indicator and one that can have wide ramifications in the development of social policy. Second, it is an indicator that can be disaggregated in a number of different ways: One can have an infant mortality figure for the nation as a whole and use it as a time series; it can be developed on a state basis, which we have now; it can be developed on a county basis. It has been used on a Census-tract basis and it can be utilized in terms of a number of such significant factors as race, income, and the age of the mother.

The third point is that infant mortality has an impact upon education: It has significant utility as a measurement of change as to what is going on not only in the health and medical care system but in a number of other areas. That's what makes it debatable because there are a lot of people who will say infant mortality is not an indicator of health per se, but an indicator of social environment including *education* of the mother and health *education* in the community.

A fourth aspect of this particular indicator is its relevance to other information. For example, it is possible to correlate data on

infant mortality with many other factors such as the extent of family planning, changes in fertility rate, spacing of children, or the number of children per mother. I like to think that the kinds of social indicators we will be developing in the education field may have that kind of usefulness. On the other hand, I recognize (and I think we all must) that because of the interrelationships that can be drawn between infant mortality and other factors, this indicator is subject to a great deal of controversy. If you use infant mortality as an indicator in a discussion with a women's group, it will be widely accepted as very important. If you mention it at a session of the American Medical Association, they will argue that it has no significant bearing on the medical delivery system. Perhaps that's the real test of a good social indicator—that different people can look at it and find different things in it that are meaningful to them in terms of social planning.

Now the fifth aspect is of even greater importance to me as a person having occupied a decision-making role in the federal government: It is an indicator that can be used to set a goal. If the infant mortality rate is X, it is perfectly possible for a reasonable person to say, "Well, my social goal is a social indicator that is 10 percent less, or 20 percent less, or 5 percent less, or 1 percent less for each year, for the next five years. And that is going to be the social goal upon which I am going to allocate scarce resources." So I think a social indicator that can be used as a measure of current development and current progress and can be used as a social goal does help a policy maker define what can be done and what should be done. Once we have taken a social indicator of any kind and made a social goal out of it, then we enter a whole new area of social policy formulation because at that particular stage, we have to analyze the impediments to reaching that goal: What are the costs of reaching that goal in terms of personnel, in terms of training programs? What are the various factors that may have some bearing on it that you can't immediately foresee? What alternatives are there? What effect will this goal have on the education system, the medical system, on parent education? A host of social policy questions begin to arise, setting in motion a discussion of social policy, which could be embodied in a social report.

I'm disappointed that nobody in the federal government or elsewhere has attempted to make another try at the social report that Dr. Olson and the rest of us had some role in developing in 1968-

69. I really think that Senator Mondale's idea of a social report is still a very good one. I would like to see somebody else try another social report to identify the kinds of social indicators that are needed to make that report meaningful. It seems to me that in the whole area of education, we are just on the threshold of seeing the kinds of social indicators that are needed for social policy throughout the country.

When I completed my role in reviewing the draft of *Toward a Social Report*, which was eventually published a few days before I was involuntarily retired in January 1969, I attempted to list various social indicators I thought were important. Subsequently, I did publish 25 of them in my introduction to *Toward a Social Report* published by the University of Michigan Press in 1970. I looked them over yesterday and I was rather shocked at the simplistic character of the 25 that I selected. I think if I were doing it today, I would try to select some additional ones.

A Quantum Leap Ahead

In conclusion, I should like to say that although you have reached only the first step in the formulation of social indicators in the education field, I see that as a quantum leap. Some day—let's say 25 years from now in the year 2000, which is not too far away—someone will look back and say, "Well, the mother of social statistics and social indicators really was a grandparent of the social report."

Afternoon Session

Measures of Educational Outcomes in Developing Countries

SELMA J. MUSHKIN
Director
Public Services Laboratory
Georgetown University

BRADLEY B. BILLINGS
Research Associate
Public Services Laboratory
Georgetown University

This paper discusses outcome measures for evaluating educational policies and programs in developing nations. The types of outcome indicators that meet the evaluation requirements range from measures of such specific achievements as functional reading to measures of competence for highly technical and professional jobs and changes in performance on the job. Some measures are especially useful in assessing projects whose immediate purpose is a certain aspect of development—for example, installing and maintaining electric transmission lines. In this case, the educational outcomes being sought are those that would be assessed primarily by the efficiency with which the electricity along the lines installed and maintained was generated throughout the area served. Other measures are especially germane to specific educational projects—for example, projects to raise the levels of knowledge about personal hygiene.

Types of Measures

A large number of different outcome measures can be useful, depending on context. Classification may be helpful. Measures of outcomes capture aspects of 1) personal development, 2) the quality of life, and 3) economic and social (including political) development. Furthermore, some indicators are functional for the immediate period of education or training, others "count" achievement on completion of schooling, and still others are yardsticks of the impact that education of today's parents will have on the

personal development of their children. In summary, the multiple educational outcomes can be displayed in a matrix that takes into account, on one hand, the different time periods (initial, secondary, and intergenerational) and, on the other, the general types of outcomes (personal, economic, and social and those indicative of education as a component of the quality of life).

Educational indicators are not always interchangeable with criteria for evaluation, which often require much greater specificity. Moreover, evaluative criteria must be sufficiently complete to encompass the objectives or goals that have been set, and the degree of accuracy has to be much greater than that for general indicators of social trends.

The decision to emphasize evaluative criteria grows out of the accelerating interest of educators and planning officials throughout the world in such questions as: Why do we want education? What is its purpose? At a meeting last year, Tanzanian officials noted: "It is now time that we look at the justification for a poor society like our own spending almost 20 percent of its government revenue on providing education for its children and young people, and. . . consider what that education should be doing." (12).

The pressure is on to evaluate educational projects and programs to determine if the intended results are being achieved. As a consequence, far more attention is being paid to educational outcomes—those intended and those achieved.

It is a long and dusty road from where it is determined that the intended purposes of education should be clarified to the point where educational outcomes are actually specified, and it is still further to the collection of the data required to assess whether, as one Tanzanian official put it, "what is intended to be achieved is being achieved."

The ruts in that road are deep. For one thing, the problem of concepts of measurements has to be resolved. It was not so long ago that the major emphasis in many countries was put on investment returns from education, with the expectation that more education would produce a higher gross national product (GNP), higher earnings, and more employment in advanced occupations at secondary or tertiary levels. Among the criteria for measurement were the GNP, the amount and percentage increase in the GNP, the level of, and changes in, earnings, the relative distribution of employment by levels of occupations, and so forth (5, 16, 17).

The lesson of that too-simplistic concept was hard for educational planners to learn. But it was learned. Experience in one nation after another emphasized that returns from education in the form of higher income and expanded GNP depend upon a combination of circumstances that are not unimportant such as capital expansion and an industrial development that creates jobs that can use the higher skills and knowledge required. Without a strong demand for workers, a highly educated labor force appears to create more public dependency rather than more output. An additional complication occurs when large numbers in the population disdain manual labor and tend to leave the rural communities of their birth in a vain effort to find work in congested cities.

In many developing nations, rates of social return on higher education have been below the going rates of return on physical capital investment. A 1970 study, for example, showed rates of return on university education below 10 percent of investment in education in Colombia, Chile, South Korea, Israel, Ethiopia, and Kenya (19).

Furthermore, experience pointed to higher returns on primary education than higher education, a result contrary to the expectations of the initial policy. A study in Kenya showed social returns on a primary education investment of 14 percent, five percentage points above that of university education (21). The findings on relative returns computed for other nations, including Puerto Rico, Venezuela, Colombia, South Korea, Israel, India, Ethiopia, Kenya, and Uganda, were similar.

A New Emphasis on Testing

Later, beginning perhaps with evaluations of school programs in 1966 and 1967, there was a turn in the road—almost a right-angle turn—when emphasis shifted from investment returns to outcome of education measured in terms of specified achievements at specified primary and secondary grade levels (1, 8). Development and application of testing instruments were encouraged in one country after another. It would be an exaggeration to imply that even the majority of nations by now have developed national standard testing instruments, but it is not an exaggeration to

note that testing is receiving far more attention than ever before. About 12-14 months ago, for example, one nation initiated a program of standardized achievement testing and planned to test all children at each school grade within a period of two or three months. Only after the hard facts about the time required to develop standardized testing instruments and the costs of application were made clear was the plan altered.

At some point along the way, the road becomes muddy. Schools—or differences among schools—seemed to account for little of the variation in tested achievement levels. This was the finding of the famous Coleman study (2). A study by the International Association for the Evaluation of Educational Achievement (IEA), headquartered in Sweden, showed similar results, but there are some differences. The latter findings seem to say that for some subject areas and in some instances, school resources do make a difference (3, 18, 20, 22).

As a consequence of these and other studies that have been made, we have learned to define educational outcome goals for planning purposes with far greater specificity and in terms that can be met. Educational outcomes, to be useful assessment measures of relative progress, must be defined in the more practical terms of whether or not they can be accomplished within the social and environmental climate in which educational services are provided. Coleman and others have found that the chief determinants of educational achievements of children are the educational level and socioeconomic status of their families. Following the evaluation already carried out, there can be no easy acceptance of the relation of school inputs to achievement scores as measures of output.

Still other studies start with disadvantaged children (children defined as disadvantaged by the characteristics of their families) and seek to determine whether differences in inputs can improve school achievements. The answer, tentatively given, appears to be "yes." Combinations of school inputs—textbooks in the case of Schiefelbein's Chilean study* and teacher behavior in the case of Blaschke's Michigan study (7)—can make a difference.

The IEA findings appear to show that there are differences in the extent to which achievements in different subjects are influenced

*Comments by Ernesto Schiefelbein during a private conversation in 1974

by family background. As one might expect, family has less to do with achievements in science and mathematics than with achievements in vocabulary skills or literature. Such findings challenge the use of achievement scores in one subject as surrogates for scores in others.

Further down the road, greater importance was attached to the notion that the way children feel about themselves and how they feel about society have much to do with whether they find it important to learn or even to take tests. Self-esteem and external-internal control, among other attitudes and attributes, have been identified as important constructs in determining a person's competence and educational performance.

The significance of the self concept and locus-of-control constructs is perhaps best analyzed from the perspective of the present cultural milieu of a society and its emerging patterns. There are many unanswered questions about the meaning of personality variables in different societies. What is the relationship, for example, between self concept and locus of control in a traditional, or structured, society, and between them and motivation to learn? One hypothesis is that in a highly traditional society, the members of lower classes may have high self-esteem (derived from the security of a class structure that has clearly defined roles and expectations of individuals) combined with an external locus-of-control orientation that breeds fatalism. The configuration perhaps might be reversed for the upper-class members of the society who have great mobility but higher self-expectations.

If self-esteem or similar attributes have an important bearing on an individual's motivation for learning, one must have information about the time required to change a person's perspective of self through alternative educational methods. To encourage learning, self-esteem may have to be developed, and the time necessary for that development would determine the minimum critical time phasing for anticipated implementation of educational plans. It certainly would have an important bearing on the time allowed for the results to flow through from individual view of self to motivation to learning as well as on the time that one must allow before plan results can be evaluated.

There is a fork in that dusty road. To the left lie modernization objectives and preparation for urban life and to the right lies improvement of life in rural communities as a goal of educational

systems. Long periods of colonization in many parts of the world, particularly in Africa, tended to fragment and denigrate existing cultural values and life patterns. People in the urban areas may have adopted European values in their educational aspirations, although some would allege such values are not always rewarding or useful to them as individuals. The emphasis on modernization that leads the Technical Institute at Kuala Lumpur to formulate *modernity* as its main educational objective is straightforward—it is industrialization, economic growth, and upgrading the education of rural youth for living in urban communities.

An extensive interview study, based on questionnaires and involving some 760 South Korean professors and more than 750 journalists, sought to relate modernization with such indicators as ability to accept changes, national consciousness, an enterprising spirit, an innovating spirit, rational judgment, and a sense of participation. Sung Chick Hong, a sociologist at the University of South Korea who carried out the study and developed adjusted measures of modernity, found that professors and journalists were predominantly "change-oriented," with tolerant, nontraditional attitudes on social change and occupational preferences (10).

Several instruments have been developed to measure modernity. Inkeles (11) and his colleagues defined modernity by a complex of traits stressing the dependence of a nation's development on the transformation of the individual. Manaster and Havinghurst (15) developed scale measures of the propensity to take risks as opposed to the propensity to seek security. One of their questions asks: Which do you prefer—a job where you are almost certain of your ability to perform well or a job where you are usually pressed to the limit of your abilities? McClelland (13, 14) and others developed the hypothesis that *need achievement* is an important element in the economic activity of individuals, communities, and nations.

The relation between self concept and economic development is critical. If the weakening of societal structure and tradition reduces self-esteem, perhaps a precondition for learning is to maintain or prop up structure in some form that is conducive to learning but not a barrier to enlarging economic activity. It may be assumed that one advantage of rural nonformal education is just this: The program is designed to facilitate learning while maintaining the orthodoxy of the former structure in things other

than education. But the objective and design of nonformal education are most important if the linking role is to be performed.

In some developing countries, a concern consistent with the self-esteem hypothesis is that the changes wrought by modernization and urbanization have been accompanied by alienation and psychological malaise. It is difficult to establish a clear cause-and-effect relationship, but it is repeatedly said that rapid changes in life styles have psychological consequences.

Political and Social Consequences

Education's outcomes in terms of political and social consequences do not go unobserved. For example, the educational objectives recommended by the Education Sector Review Commission of Ethiopia were mainly concerned with social and political factors (6, 9). A major recommendation that was made by the Commission on Primary Education for the Conference on Education and Scientific and Technical Training in Nairobi was that "the African government should insure that the primary education system . . . contribute to the strengthening of national unity . . . bring about the social and cultural integration of children in the community and act as factors of change and of economic and social development." (4).

McClelland also has shown that training can produce the type of modernity that is required. In a controlled experiment, he demonstrated that need-achievement scores can be increased. Participants trained for modernity were found to be more active; they actually started new ventures, mobilized increased amounts of labor and capital, and tended to be responsible for a relatively larger percentage of growth in a firm's gross income than control group members.

Educational objectives linked to change and modernity have been widely accepted in many countries. As one observer indicated, the faith that is placed in education as an agent of change is of such far-reaching importance, in fact, that much greater attention has centered on developing and adapting education to meet the needs of rapid development than on meeting needs of rural living. Education is the key that unlocks the door to modernization, and modernization and rurality have been mostly opposite

indicators. For example, the percentages of population engaged in farming are contrasted with the percentages in manufacturing or in the technical and professional trades. The higher the percentage in the latter, the greater the modernization. Or the percentage of the population in rural areas is contrasted with that in urban areas, with the concept that the higher the urban percentage, the more modern the country.

More recent moves toward rural education take a somewhat different view of the critical requirements of a developing nation. Industrialization is considered a complicated process that depends on changes in those institutional structures that play a primary role in providing opportunities and incentives for change in the individual. Unless the institutional structure is conducive to development of the urban community and to employment, the argument goes, it is well to foster rural development and to keep the population back on the farm. Thus, the stress is on rural education, on making the most of development in the rural community, and on putting those with knowledge and skill to work to foster rural development. Ethiopia, for example, in launching its rural development campaign, gave as a stated objective the bridging of the wide intellectual and attitudinal gaps between the educated and the uneducated by making one learn from the other through discussion and hard work.

The main criteria in terms of the objectives sought are measurements of rural progress in terms of personal development, quality of life, and agricultural output.

Some of the measures of rural skill and knowledge that are important are these objectively measured indexes:

- Number and percent of water resources safe for drinking
- Percent of water resources used for both bathing and drinking
- Percent of women of childbearing age who have a choice about bearing children
- Percent of persons who wash hands before preparing food and eating
- Percent of persons who have such avoidable diseases as hookworm and tapeworm
- Number and percent of persons immunized against disease
- Number and percent of infants who die at birth or within first week

- Amount of exports from the communities and relation of exports to imports
- Number of teaching units available in the community
- Number of persons enrolled in educational offerings and the percent of relevant age groups represented
- Number and percent of children and of adults who achieve basic literacy and knowledge of arithmetic
- Number and percent of adults who regard blue collar jobs and white collar jobs as equal in esteem
- Number and percent of adults who apply newly learned information about nutrition to family feeding
- Number and percent of adults who maintain their dwelling units with the materials available in the community
- Number and percent of adults who work on community property improvement projects; percent of time spent

There are also process variables that spell out ways to enlarge agricultural output, improve facilities for education or health, or provide access to markets for handicrafts and agricultural products. Among such process variables are:

- Number and percent of farm units using new tools
- Number and percent using fertilizer in production of foodstuffs
- Kilometers of roads under construction as a percent of all roadways
- Number of springs for drinking water that are fenced off from animals

It is possible to view rural economic, social, and personal development as end purposes of education. Whether nations that are poor in resources and that lack physical capital can significantly improve the well-being of their populations by higher farm and crafts output remains to be explored. There are few who would dispute, however, that more farm output per capita has to be accompanied in many nations by altered perspectives about the size of families and about childbearing, lest basic Malthusian principles operate to wipe out any gains. Therefore, some components of individual development are important to the achievement of the economic development goals. The question of how far to proceed along the path to personal growth is a difficult one that different societies undoubtedly are answering in a number of different

ways. Among the choices are these:

- To restrict educational offerings to those that will enhance rural development such as education for family planning, greater use of new agricultural tools, fertilizers, and so on
- To include in education not only the restricted offerings but those of a cultural type that will emphasize literacy and history and those that provide traditional education of young children
- To broaden educational offerings for greater equity; for example, to facilitate occupational mobility along with expansion of community services and opportunities for a broader spectrum of occupations in rural areas

Tanzania, in its formulation of rural education for children, chose the second of these options. It undertook to assess, for example, the extent to which each child by "standard seven"

- Achieves mastery of defined subjects—kiswahili, mathematics, social studies, and domestic science (or agriculture or crafts)
- Values equally all subjects of the curriculum
- Regards primary education as terminal

In Ethiopia, the choice has been different. The constraints of a seventh-grade education have not been accepted. Rather, the emphasis was put upon evening out the discrepancies between rich and poor and between educated and uneducated. The basic education program in fundamental literacy was designed, among other things, to lay the foundations for the acquisition of future skills.

In many developing countries, nonformal education is the process currently in vogue to achieve the kind of learning that is being sought. Much of the machinery currently being developed for evaluation is designed to assess achievement in subject matter areas and to improve the primary and secondary school-leaving examinations so that candidates may be channeled into the next sequence of formal education. In the Caribbean, in the Philippines, and in Pakistan and Somalia, to mention just a few places, many resources are going into testing instruments for school-leaving examinations.

But in many ways, resources for measuring outcomes of education are needed most for nonformal education. It is in that context that passing from one grade to another surely cannot be a

criterion for learning. The many educational indicators of an input or even a process type do not provide the required measure of output. Some of these familiar indicators are:

- Number and percent enrolled in school, by age and status of children
- Number and percent of school graduations
- Number and percent of school dropouts
- Expenditures per child at each schooling level
- Size of class or teaching unit
- Percent of teachers who are "trained"

If flexibility of movement is to be maintained between types of education—nonformal to formal and vice versa—a way has to be found for judging competency not by time spent in school or even grades completed, but by learning acquired—the knowledge, skills, and attitudes demanded for advancement in the sequence of study. Also, perhaps to a greater extent than now, criteria of relevance will become necessary. Whatever the tests of criteria of assessment, they should be derived from the actual requirements of that sequence. And, if professional or technical competency is the goal, the specific content of that competency has to be defined so that those who seek to enter from a nontraditional track can do so. To illustrate, those who perform effectively as sanitarians or midwives in rural areas should be given the opportunity, if they wish, to advance to medical training. Those who learn well and apply skills of husbandry and nutrition should, if they so decide, be helped to continue their education and training.

Whether or not competency is redefined, it is critical to flexible movement between nonformal and formal education to find ways to determine knowledge, skills, and attitudes acquired and to arrange that the sequence of formal learning can be adapted to entrance and exit from nonformal education. Perhaps, in this effort, more research is required on the validity of existing testing instruments as predictors of success in formal schooling and on the job or in performing the work.

On the long and dusty road toward development of measures of educational outcomes, some attention has been given to understanding the fundamental requirements of such occupational skills as those of an electrician, a carpenter, or a welder. Requirements are not only being defined, they are being specified in terms

of different technologies: that of an advanced nation and those of nations at various stages of development in which the occupational skills are being applied. The effort by the UN's International Labour Office (ILO) along these lines has not been linked to the nonformal educational policies fostered by UNESCO and others. But in rural communities, the idea is applicable to development of measures of knowledge and skills, including weaving, pottery, husbandry, baking, road construction, and house construction.

The number and proportion of dropouts from formal schooling—indicators often used to judge education—have little meaning in the context of a flexible system in which individuals move between programs. It is the application of the knowledge and skills acquired while learning that is important to assess. Dropping out of formal school is sometimes termed "wastage." In the context of nonformal education, it may mean the achievement of greater efficiency.

Greater attention is being given by the developing nations to evaluative criteria by which education can be assessed. Out of this new effort comes new attention to outcomes from education. And, in turn, assessment instruments and new indicators are being formulated in support of evaluative efforts and program decisions to achieve greater rural development and equity among groups in a nation.

References

1. Coleman, J.S. The evaluation of equality of educational opportunity. In Frederick Mosteller & Daniel P. Moynihan (Eds.), *On equality of educational opportunity*. New York: Vintage Books, 1972. Pp. 146-167.
2. Coleman, J.S., *et al. Equality of educational opportunity*. Washington, D.C.: U.S. Department of Health, Education, and Welfare, 1966.
3. Comber, L.C., & Keeves, J.P. *Science education in nineteen countries*. New York: Wiley-Halstead Press, 1973.
4. Commission on Primary Education for the Conference on Education and Scientific and Technical Training. Nairobi, Kenya. 1968.
5. Denison, E. Measuring the contribution of education to economic growth. In John Vaizey (Ed.), *The residual factor and economic growth*. Paris: Organisation for Economic Co-operation and Development, 1964.
6. Education: challenge to the nation. Report of the Education Sector Review. Addis Ababa, Ethiopia, August 1972.
7. Education Turnkey Systems, Inc. *Summary cost effectiveness study of compensatory education*. Lansing, Mich.: Michigan Department of Education, March 1975.
8. Evans, J.W. Head start: comments on the criticisms. In Francis G. Caro (Ed.), *Readings in evaluation research*. New York: Russell Sage Foundation, 1971. Pp. 401-407.
9. Fassil, G. Kiros, *et al.* Financing of education. Report of Education Sector Review 5, Addis Ababa, Ethiopia, 1972.
10. Hong, Sung Chick. *The intellectual and modernization*. Seoul, Korea: Research Institute, Korea University, 1967.
11. Inkeles, A. Making men modern: or the causes and consequences of individual change in six developing countries. *American Journal of Sociology*, 1969, 72, No. 2, 208-225.
12. International Institute for Educational Planning. Education for self-reliance. Report of the African Regional Seminar on Educational Evaluation, Dar es Salaam, Tanzania, spring 1975. Paris: International Institute for Educational Planning, May 1975. P. 10.
13. McClelland, David. *The achieving society*. Princeton, N.J.: D. Van Nostrand Company, Inc., 1961.

14. McClelland, David, & Winter, David G. *Motivating economic achievement.* New York: The Free Press, 1969.

15. Manaster, Guy J., & Havighurst, Robert J. *Cross-national research, socio-psychological methods and problems.* Boston: Houghton Mifflin, 1972.

16. Organisation for Economic Co-operation and Development. *Financing of education for economic growth.* Paris: Organisation for Economic Co-operation and Development, 1966.

17. Organisation for Economic Co-operation and Development. Policy conference on economic growth and investment in education, Washington 1961. Paris: Organisation for Economic Co-operation and Development, 1966.

18. Platt, William J. Policy-making and international studies. Paper presented at Conference on Educational Achievement held at Boston, Mass.: Harvard University, November 1973.

19. Psacharaopoulos, G. *Returns to education.* London School of Economics, Higher Education Research Unit. Studies on Education. London: Elzevier Scientific Publishing Company, 1973.

20. Purves, Alan C. *Literature education in ten countries.* New York: Wiley-Halstead Press, 1973.

21. Thias, Hans Heinrich, & Carnoy, Martin. *Cost benefit analysis of education: a cost study of Kenya.* Washington, D.C.: International Bank for Reconstruction and Development, 1972.

22. Thorndike, Robert L. *Reading comprehension education in fifteen countries.* New York: Wiley-Halstead Press, 1973.

Discussion: Priority Research Agenda on Educational Indicators

Discussion

ELEANOR BERNERT SHELDON

I'm astonished at what goes under the umbrella of social indicators and what goes under the umbrella of educational indicators. (I'm not commenting on things that Dr. Withey has just brought out because I do think that they are quite relevant.)

I have a feeling that those indicators we will be developing for an understanding of our educational system—and I do urge you to view it as a system as opposed to a set of disparate activities—those indicators we will be capable of developing for describing and analyzing our own educational system in this country, and how it relates to other systems in this country, will be quite different from those measures we'll be using on a cross-national basis. And I would suspect we should not immediately try to design measures that will both describe what we are doing and have applicability elsewhere.

So I urge, first of all, that those of you who are concerned with examining the educational system in this country do not for the moment worry about designing measures that will be cross-nationally applicable. I think that one of the first attempts to examine the educational system was very admirably made on this platform this morning by my colleague, Mr. Gooler. I think that he has provided us with some kind of an approach to looking at the educational system. I do not believe that it is a perfect approach and I'm sure he doesn't either. It has been placed before you as a reference point. Why don't you try organizing some of the currently available data? Heaven knows, there is a great deal of that! There are time-series data. (And please bear in mind that we are talking about change.) I think that if once you started organizing data either along the line of Mr. Gooler's proposals or any other that you have and viewing education as a system within itself and then examine the relationships within that system and

then the relationships of that system to other systems, you would come up with a wide range of modifications, refinements, and new starts on conceptualizing what it is you're after.

I think one of the first things we want to do is see what data we have that describes what we are. I think we have a start here in our *Condition of Education*—I'm sure it can be revised and improved. I think we should try to find out what we have before we begin designating what more we want to know. Let's see where we are in describing a system that we can at least grossly define and look at some of those data before we start rushing into either cross-national work or a lot of new data-collection enterprises. I think that it's very important that we become fully acquainted with some of the ongoing studies that both the government and some of our private and university agencies are engaged in. I think the National Center for Education Statistics may very well prove to be an important locale for our work in educational indicators. The Center has access to a wide range of government materials and a talented staff who can begin organizing those materials around any framework that initially sounds plausible. I think they are the important custodians and (I hope) considerate thinkers about a very important longitudinal study which has been in the field for almost five years now. I think there are very interesting data that will emerge from that arena. Also, as I said earlier this morning, we should be criticizing or getting ready to criticize not only the National Center work but also the educational assessment work.

I think that there is a confusion about the differences between social indicators and how one evaluates social intervention programs. And I would hate to see this group go away without getting that cleared up. Social indicators, as time-series measurements that can be disaggregated as far as our samples will allow us to disaggregate them, are not capable of evaluating social program interventions. The logic of the two approaches is different. We are not operating in a closed system, as the economy presumably was. If you institute a program of vocational guidance, for example, no amount of time-series data is going to tell you whether or not that program was successful or what was successful about it or what made it succeed or what made it fail. If we could go away clearing up only one confusion, it would be a very worthwhile day—namely, time-series statistical data are capable

of describing and analyzing some systems. They are *not* capable of evaluating the output of program interventions.

For those of you who are going to work in the arena of educational indicators, I urge you not to worry, not to debate, and not to argue about what is and is not appropriate. I think the most important thing is for people to get to work, to get their nails dirty with the data, become the Dickens type of data grubbers that Selma has referred to. No amount of abstract discussion as to what is appropriate and what is not appropriate and for what purposes is going to move the enterprise any further than it is right here today.

Discussion

Dennis D. Gooler

I became interested in social indicators for two reasons: First, I was interested in educational evaluation—that is, how one determines the worth or value of an educational program; second, I was interested in something called policy analysis—that is, how we as a public decide what we will feature and where we will put our resources. These two concerns seem to come together in what are called social or educational indicators. I'd like to suggest several areas that I think need some more thinking. (It may be a little grandiose to call it research.)

The first addresses the question: Why have indicators anyway? I think Eleanor's suggestion that we not spend a lot of time navelgazing but try instead to do something about indicators is right on.

I think it would be helpful to examine the possible consequences of seeing indicators both as descriptions of trends and of status at different points in time and as particular and unique input to policy making and decision making. I think that there are different ramifications that accompany these two broad views of educational indicators. We ought to spend time trying to spell out the possible consequences of holding each of those points of view.

We should try to get a little better handle on why it is we're trying to come up with these indicators anyway. Ideally, I would like to see us be able to determine the relationship of processes and outcomes of education, so that we might have a somewhat better predictive model of what happens when we try to do certain things. Another ideal would be to establish clearly and definitively the relationships between education and other social variables, such as crime, poverty, and so forth.

But those are ideal, and I don't think we'll make a lot of progress at that global level. Instead, I suggest that we think about trying

to refine existing data or develop new kinds of indicators, possibly in one of the areas I've mentioned, such as the impact of education. It's easy to talk about what ought to constitute a good information system with respect to any of these categories; it's quite another thing to actually pull data together or to conceptualize what additional data are needed.

I think we also need to try to test out the feasibility of any of these indicator systems. If we propose some data-collection efforts which are simply impossible to implement, we haven't gotten very far. We ought to check out the practicality of producing and interpreting certain kinds of data and, having done so, to assess the implications of having those data—to see how congressmen or parents or teachers or professors or whoever interpret these data, what kinds of decisions they make, how the data change their own view of the world, and so forth. I think we need to focus on a small number of these things, try them out, determine their feasibility, and figure out what we've got once we've got a good system of educational indicators.

All of this assumes that we are able to articulate bases of policy, that we're able to figure out the ingredients that go into making decisions about resource allocations, goals, and so forth. These decisions are clearly not always made rationally, nor should they necessarily be. I think that if we are going to *advocate* the use of indicators, we have to understand a great deal more about how to collect these data, how to interpret them, and what happens once you have educational indicators.

Discussion

Stephen B. Withey

We have indicators that tell us a great deal about the educational establishment—about its size, its cost, it indebtedness, its activity, its characteristics, its participants, and so on. And I suppose we need those badly and we could improve them. And if we could get simpler indices with larger meaning, that would be progress. We obviously have several measures of accomplishment in regular and irregular testing programs, and they essentially measure the content of the curriculum. If one is interested in curriculum, those obviously should be good indicators and they should be improved and verified as the curriculum changes.

We also have some measures of the achievements of alumni. Occupational, economic, and several other criteria of success tell you something about the product of education; and if those can be improved, they would be helpful.

There are two or three other points of view I would like to give some priority to. I think we do a fair job now of assessing the primary purposes and consequences of education. But we don't do as good a job of assessing the unintended, or secondary, consequences of education, both advantages and drawbacks. It's hard—perhaps the roughest measurement problem one could pick—but it is so important that I think we should try. In that sense, I would give it priority of interest, although I don't give it priority of promise. We need to know how the educational experience really meshes with adult life. In a sense, that is what education is all about.

One might consider some index of the basic or minimum skills (literacy, computation—that kind of ability) that education should give adult life. It's not the whole story, but it seems to me it's of great importance. We should have some measure of salable skills. We should also have some measures of adaptive skills for

living in our society. They're harder to measure, but I think they are important. I think we should have something on appreciative skills and know-how. How do we appreciate life and people and the arts?

Finally, I think we should be studying something about the values and issues that we are sensitive to as we finish formal education. I mention those not in the sense of goals. I don't have time to explain what I do mean by those, but let me just give you a hint. We often collect data on how people evaluate events. We call such data "attitudes" toward people, activities, or objects. There is considerable data to indicate that we can get more predictive power on behavior and better understanding of it if, instead of measuring how we feel about the event, we ask people what their beliefs and assumptions are about how it came to be and what its consequences are. We may do even a little better if we ask about the other optional courses the event might have taken. These perceptions of the contexts of events, which may be heavily influenced by education, may be assessable and may be revealing indicators of how people are interpreting their life experiences.

Discussion

MANCUR OLSON

I forgot to say earlier that I'm not a specialist on education. Perhaps you have learned that by now. I must emphasize that my aspiration has been to bring some principles from my craft of economics to bear, in a general way, on the kinds of concerns of people at this conference, and since you are experts on education or educational measurement, to leave it to you to improve the way in which these ideas are applied to education.

I would also like to supplement what I said in my talk on the importance of multiple causation in producing education and many other things as well. Of course, there is nothing new in saying that how much children learn depends not only on the schools but also on the home environment, the neighborhood, television, and the like and that social indicators describing other aspects of reality are normally also influenced not only by some single government agency but also by diverse social and economic forces. This is a point that has been widely understood for some time and which was dramatically underlined in the Coleman report. Thus, I mentioned multiple causation in my talk and have been emphasizing it ever since I began to think about social indicators, not because of any illusion that I am being original, but rather because it interacts with the public-good measurement problem I was talking about in a way which makes that problem more important and difficult.

For the purpose of the present argument, let's accept what I said before lunch about the special difficulty of measuring the outputs of public goods. Then let us ask what connection there is between my argument about public goods and the problem of finding out how much each of the relevant independent or causal variables influences a particular social indicator or dependent variable.

When we ask about this connection, it immediately becomes clear that, even if we had *perfect* measures of every independent variable that influenced some social indicator except one that was a public good (a "public intermediate good"), we would not, in general, be able correctly to estimate (or perhaps even to specify) the relationship between *any* of the independent variables and the dependent variable. The absence of a measurement of the public good, or the unknown bias in its measurement, will mean either that we have a "specification error" in our estimating equation or a bias in the measurement of one of our independent variables. In either case, our estimates of the relationship between *other* independent variables—the ones that do not measure the output of public goods—and the social indicator are likely to be in error. The general difficulty of explaining the causes of changes in social indicators, even when particular causes aren't public goods, is due in substantial part, then, to the special problem of measuring the outputs and production functions of public goods that I have been talking about. This difficulty of estimating the relationship between independent variables and social indicators, in turn, means that we cannot determine the outputs in production functions of public goods as residuals.

What has just been said may be made clearer by means of an example. Police protection is a public good and, by my argument, the output of the police department is not likely to be measured, at least not accurately, and any judgments about how it is changing are likely to be biased by the observer's ideology. It is, however, possible to get some social indicators on crime, such as crime rates obtained from victim surveys. If we haven't got meaningful data on the output of the police, though, we can't trust the regressions we might run on the relationship between income, unemployment, street lighting, family stability, or what have you, and any crime rates.

These problems are not as difficult in the area of education because education is partly a private good, but even here they have proved to be difficult enough.

Okay, if it's so difficult, why bother at all? One reason we should bother came up in my talk this morning: A little bit of information can have tremendous incentive effects and arthritis-reducing effects in school systems and in other public bureaucracies.

Another reason to bother with social indicators and output

measurement is that there is no other way to get information about social production functions or the causes of social change that is generally any easier. Some will disagree with me and say that there is an easier way: the way of controlled experiments. In view of the statistical and conceptual difficulties of estimating production functions with the aid of social indicators and other statistics, they ask, why not use controlled experiments? The answer is that we should indeed use social experiments; they can sometimes advance our understanding considerably, as in the case of that famous experiment with the negative income tax in New Jersey.

Note, however, that if public goods are at issue, an effort to set up an experiment brings us back into the same difficulties we have been discussing. This is because of the fundamental logic of the situation. If a public good is being produced, it follows that it will be impossible, or at any rate very costly, to provide the good for the experimental group without providing it for everyone: This impossibility or costliness follows from a defining characteristic of public goods. With public goods, in other words, the experimental group may have to be as big as the group that uses the good in real life, so that the distinction between experimental cases and real or practical cases is no longer meaningful. Added to this is, of course, the fact that where human beings are concerned, there are both moral and practical limits to what can be tried, even when the experiment offers a prospect of useful information.

Thus, whatever else needs doing, there is also a need, after facing up to the difficulties, to attempt to get what insights one can from statistics on real situations—on the "natural experiments" that take place at different times and places. We must gather more social indicators, try to get better measures of the independent variables that are not public goods, and attempt to get the best information we can on the outputs and production functions of public goods.

This data and information, in turn, should be, to the greatest extent possible, in a form which enables the researcher to study the relationships among the relevant variables. For historical reasons, the social indicators produced by the National Assessment, for example, have not in the past been in a form that facilitated an analysis of the effects of particular schools or kinds of schooling on those whose performance is being assessed. When I

last looked into the matter, at least, the National Assessment could not publish data on what schools those who were tested had attended.

I am sorry to go into this slightly technical matter of multiple causation at such length, but this is a matter about which there has been a bit of confusion in some of the critical literature on social indicators. If these confusions are not cleared up, they might distract or delay experts in testing and measurement from the urgent business of using the social indicators they produce to help us get better insights into the ways the educational process can be made more effective.

Discussion

WILBUR J. COHEN

I have thought of some 10 or 20 different areas in which indicators might be helpful in developing some new social policies for the next 20 years and I want to discuss two or three of them. The first would be early childhood education. I think that we are entering a period of reexamination of what resources in our society we should, could, and will put into early childhood education. This is not merely because child psychologists have indicated that they believe certain changes in achievement, potential, motivation, creativity, or whatever we want to change might be important (and all those are debatable), but simply because with more women going to work, we are faced with a need for some kind of child care ranging all the way from custodial and day care to early childhood development of the most sophisticated and complicated type. And the question of whether the nation is going to put 5, 10, 15, or 20 billion dollars more of new money into that area raises various kinds of questions from economists and others about whether those incremental dollars, in terms of scarce resources, are going to be wisely spent or not and what value we are going to get from them. We may be faced with those kinds of decisions to make during the next decade or two with very few indicators of achievement, results, or output. We may have a lot of input information for early childhood education; but the kind of output data we will have deserves a good deal of attention. You have all sorts of problems about differential impact on low-income people. It's very probable at the present time that welfare recipients are probably getting more day-care funds for their children in terms of public resources than people at upper-income levels are getting now. So that the whole relationship between income levels and social and educational status in early childhood fields seems to me to be very ripe now for someone to try to see what kinds of indi-

cators and what kind of information we can get that will bear on that matter.

Second, in the area of minority and ethnic differences in this country, there are many issues about which there really is very, very little information at the present time. Is it possible to develop some continuing series of indicators in that area that would be reasonably reliable and useful? I have my doubts about being able to do it, but if you are asking me a question about what needs to be done in research for the future, I would say that represents a very important area.

The third area in which indicators will be important in the future is in adult and continuing education. Here I think indicators can be developed, particularly for different age levels of people in different kinds of learning environments. I think that is extremely important because we have got to have some educational indicators that are outside the normal schooling process. It is extremely important to develop these kinds of indicators so that we are not relying entirely upon those that are conventional in terms of K-12 or higher education or postsecondary kinds of environment.

Another one concerns the question of whether we can measure the improvement in the quality of teacher competence. We're entering an era of concern about teacher centers and professional growth and development centers in relation to the attitude of organized teacher associations about in-service training and professional training that is outside the normal concept of institutions of learning. That may be good or bad. That may be desirable or undesirable, but it's not going to be within the normal institutional framework of a university or a college or a community college or an educational system. We're going to have to try to develop some indicators of whether our investment to improve the *quality* of teaching goes against our concern about expanding the *quantity* of teachers and whether that investment is really something the taxpayer and others think is important.

I will conclude with two or three generalized points about the new kinds of indicators. I want to underline what Dr. Hodgkinson said this morning about the need for our having data—any indicators—that can be usefully broken down on a state and local and Census-tract basis. My own long years of experience lead me to believe that in our kind of governmental structure with the tremendously important role of states and localities in the educa-

tional system, simply having aggregate national data is not useful in terms of the implementation of social policy when you have such great variance in the United States. Therefore, I feel that we should select those kinds of indicators and those kinds of measurements that would be useful in our research. Breaking down any measurement to the smallest unit possible and comparing it with other units leads not only to a further discussion of policy but, as Mancur Olson pointed out, perhaps it would ultimately lead to a greater degree of social experimentation as well. In the future, with possibly slower economic growth, if that is a valid assumption, and with other constraints upon the public sector's role in inaugurating new programs, we may have to engage in a kind of social experimentation in small units.

Senator Ribicoff suggested several years ago that instead of our starting a minimum-income-guarantee program across the United States, we ought to take a couple of states or a couple of communities and try it out and see what the problems are. Now, as Dr. Olson said, that idea is fraught with difficulties because you can't solve problems by drawing geographical lines. On the other hand, I think the idea of experimenting in small units is more likely to happen because people are going to be concerned about the tremendous investment of resources involved in doubtful areas of social policies. And so it seems to me that while we should have no limitations on the kinds of social indicators that people can create, I would hope that we would be able, in the course of time, to define some of those that have more consequential bearing on eliminating the extent of variance and inequality in the American structure, that would lead toward the development of social policy and social change, and third, that would lead toward the possibility of a greater degree of social experimentation before we finally decide where we are going to put our scarce resources.

Discussion

Selma J. Mushkin

I'd like to talk about what needs to be done in research and take an approach that is somewhat different from Eleanor Sheldon's. I think there are some instances in which we do need much more research. I think we need, for the developing countries, a new way of looking at modernity, *rural modernity*. We need to know whether through rural development we can achieve more equitable national development. We need to know not only how we measure it, but, if it can be measured, how to create it. And we need to know whether we can create it in a relatively short period of time.

I don't think the items now used to measure self concept and external/internal control either in the U.S. Office of Education high school longitudinal study or the old Coleman study are sufficient for the developing countries or for our own. It's true that a great deal of work has been done on self-esteem and control both in defining constructs and developing instruments. But I do think we need to know more about how best to get at these constructs and whether the particular items that we now have in the longitudinal study, for example, are adequate. Let me say that the self-esteem and external/internal control items in the Coleman report had more explanatory power than family background in terms of educational achievement.

We need to know more about the interrelationships between self-esteem and control. I offered one hypothesis here for the developing countries—namely, that as you maintain a rigid social structure, you do, in fact, have a good deal of self-esteem possible in the lowest classes. And if you start breaking that class structure, it can be very disruptive. I think we need to know something about how to modify these instruments in terms of various cultural patterns, both in our country as well as different parts of the developing world. And we certainly need to know whether you

can have a training program that can be carried out effectively within a short period of time.

Also, perhaps we need in the United States and certainly in the developing world a structured way of measuring work skills. Not everywhere in the United States do you have to have equally sophisticated skills to do the same kind of thing, because the materials that you work with are different. Certainly in developing countries you don't really need the same kinds of skills for construction as are needed elsewhere. In rural communities in developing countries, you do not need the skills required to build skyscrapers. I should think we want to be able to train people specifically for the level of skill required where they are and to test for those skill levels, and at the same time, give them a possible way to move up. Now maybe those things are not social indicators but they are really what you need to evaluate as criteria of a work-skill development program.